Journal of
Experimental Psychology

DAVID A. GRANT, Editor
Department of Psychology, University of Wisconsin
Madison, Wisconsin 53706

JAMES E. DEESE, Associate Editor
Johns Hopkins University

HOWARD H. KENDLER, Associate Editor
University of California, Santa Barbara

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
FRED ATTNEAVE, *University of Oregon*
LYLE E. BOURNE, JR., *University of Colorado*
ROBERT M. BOYNTON, *University of Rochester*
GEORGE E. BRIGGS, *Ohio State University*
CLETUS J. BURKE, *Alameda County State College*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *Princeton University*
J. J. GIBSON, *Cornell University*
JULIAN HOCHBERG, *Cornell University*
ARTHUR L. IRION, *Tulane University*
GREGORY A. KIMBLE, *Duke University*
HERSCHEL W. LEIBOWITZ, *Pennsylvania State University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
BENNET B. MURDOCK, JR., *University of Vermont*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *University of Iowa*
HAROLD W. STEVENSON, *University of Minnesota*
DELOS D. WICKENS, *Ohio State University*

HELEN ORR
Managing Editor

RUTH P. JAGOLINZER
Editorial Assistant

ABETH S. REED
Advertising Manager

Volume 66, 1963

PUBLISHED MONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA. 17604
AND 1333 SIXTEENTH ST. N. W., WASHINGTON, D. C. 20036

Second-class postage paid at Lancaster, Pa.

CONTENTS OF VOLUME 66

ALLEN, C. K. See WICKENS, D. D.
ANDERSON, J. E. See REYNOLDS, W. F.
ARONFREED, J. See LEWIS, M.
ASDOURIAN, D. See SALTZ, E.
BAIRD, J. C. **Retinal and Assumed Size Cues as Determinants of Size and Distance Perception** 155
BECK, S. B. **Eyelid Conditioning as a Function of CS Intensity, UCS Intensity, and Manifest Anxiety Scale Score** 429
BELL, C. R., AND PROVINS, K. A. **Relation between Physiological Responses to Environmental Heat and Time Judgments** 572
BERGUM, B. O., AND LEHR, D. J. **End-Spurt in Vigilance** 383
BERKLEY, M. A. **Discrimination of Rewards as a Function of Contrast in Reward Stimuli** .. 371
BERLYNE, D. E., CRAW, M. A., SALAPATEK, P. H., AND LEWIS, J. L. **Novelty, Complexity, Incongruity, Extrinsic Motivation, and the GSR** 560
BESCH, N. F. See REYNOLDS, W. F.
BEVAN, W., AND PRITCHARD, J. F. **Effect of "Subliminal" Tones upon the Judgment of Loudness** 23
BILODEAU, I. McD. See SAUFLEY, W. H., JR.
BINDER, A. **Effects of Altered Frequencies upon Recognition Responses** 553
BLATTNER, K. C. See PETERSON, M. J.
BOBLITT, W. E. See BRACKBILL, Y.
BOGARTZ, R. S., AND CARTERETTE, E. C. **Free Recall of Redundant Strings of Symbols by Children** 399
BOUSFIELD, W. A., AND COWAN, T. M. **Conditioning of Motor and Verbal Responses to Nonverbal Stimuli** 47
BOWER, G., AND TRABASSO, T. **Reversals Prior to Solution in Concept Identification** 409
BRACKBILL, Y., BOBLITT, W. E., DAVLIN, D., AND WAGNER, J. E. **Amplitude of Response and the Delay-Retention Effect** 57
BRACKBILL, Y. See LITTLE, K. B.
BRALEY, L. S., AND JOHNSON, D. M. **Novelty Effects in Cue Acquisition and Utilization** 421
BROADBENT, L. See GOTTSDANKER, R.
[illegible]DEN, M. P., AND RAINEY, C. A. **Left-Right Differences in Tachistoscopic Recognition** ... 568
[illegible]DI, E. J., AND SPIVEY, J. E. **Effect of Goal-Box Similarity on the Aftereffect of Non-[illegible]nforcement and Resistance to Extinction** 461
CAR[illegible]N, J. B. **Effect of Amount and Distribution of Inspection Time and Length of Decay Interval on Kinesthetic Aftereffect** 377
CARTERETTE, E. C. **A Replication of Free Recall and Ordering of Trigrams** 311
CARTERETTE, E. C. See BOGARTZ, R. S.
CHERNIKOFF, R., AND LEMAY, M. **Effect of Various Display-Control Configurations on Tracking with Identical and Different Coordinate Dynamics** 95
CLAYTON, K. N. **Reversal Performance by Rats Following Overlearning with and without Irrelevant Stimuli** 255
CLAYTON, K. N. See ROBINSON, A. W.
COHEN, B. H. **Recall of Categorized Word Lists** 227
COHEN, B. H., AND HUT, P. A. **Learning of Responses to Stimulus Classes and to Specific Stimuli** 274
COLE, C. See MARX, M. H.
COWAN, T. M. See BOUSFIELD, W. A.
CRAW, M. A. See BERLYNE, D. E.
CREAMER, L. R. **Event Uncertainty, Psychological Refractory Period, and Human Data Processing** 187
DALLETT, K. M. **Practice Effects in Free and Ordered Recall** 65
DAVLIN, D. See BRACKBILL, Y.
DORNBUSH, R. L. See WINNICK, W. A.
DOROZ, L. See ERIKSEN, C. W.
DOUGHERTY, D. See MARX, M. H.
EBEL, H. C. See PROKASY, W. F.
EBENHOLTZ, S. M. **Serial Learning: Position Learning and Sequential Associations** 353
EPSTEIN, W. **Attitudes of Judgment and the Size-Distance Invariance Hypothesis** 78
ERICKSON, R. L. **Relational Isolation as a Means of Producing the von Restorff Effect in Paired-Associate Learning** 111
ERIKSEN, C. W., AND DOROZ, L. **Role of Awareness in Learning and Use of Correlated Extraneous Cues on Perceptual Tasks** 601

ERIKSEN, C. W., AND HOFFMAN, M. **Form Recognition at Brief Durations as a Function of Adapting Field and Interval between Stimulations** 485

ERLEBACHER, A. **Reversal Learning in Rats as a Function of Percentage of Reinforcement and Degree of Learning** 84

ERLICK, D. E. **Effects of Grouping of Stimuli on the Perception of Relative Frequency** 314

ERLICK, D. E. **Effects of Method of Displaying Categories on the Perception of Relative Frequency** 316

FEHRER, E., AND GANCHROW, D. **Effects of Exposure Variables on Figural Aftereffects under Tachistoscopic Presentation** 506

FLEISHMAN, E. A., AND RICH, S. **Role of Kinesthetic and Spatial-Visual Abilities in Perceptual-Motor Learning** 6

FORRIN, B. See MORIN, R. E.

FORT, J. G. See MYERS, J. L.

FRIEDMAN, S. M. See HARCUM, E. R.

GANCHROW, D. See FEHRER, E.

GOLD, C. See TULVING, E.

GOSS, A. E. See MUSGRAVE, B. S.

GOTTSDANKER, R., BROADBENT, L., AND VAN SANT, C. **Reaction Time to Single and to First Signals** 163

GREENO, J. G., AND LABERGE, D. **Sequential Dependencies and Nonreinforcement in Probability Learning** 547

GRICE, G. R., AND HUNTER, J. J. **Response Mediation of the Conditioned Eyelid Response** .. 338

GRICE, G. R. See KOPPMAN, J. W.

GRINGS, W. W., AND LOCKHART, R. A. **Effects of "Anxiety-Lessening" Instructions and Differential Set Development on the Extinction of GSR** 292

GUIRAO, M. See STEVENS, S. S.

HAMMERTON, M. **Retention of Learning in a Difficult Tracking Task** 108

HAMMERTON, M. **Transfer of Training from a Simulated to a Real Control Situation** 450

HARCUM, E. R., AND FRIEDMAN, S. M. **Reproduction of Binary Visual Patterns Having Different Element-Presentation Sequences** 300

HEASLET, G. See MARTIN, J. G.

HILL, F. A. See WICKENS, D. D.

HODGSON, D. A. See WEISS, W.

HOFFMAN, M. See ERIKSEN, C. W.

HOFFMAN, P. J. See PRADHAN, P. L.

HOM, G. See MARTIN, J. G.

HOVING, K. L. **Influence of Type of Discrimination Training on Generalization** 514

HUNTER, J. J. See GRICE, G. R.

HUT, P. A. See COHEN, B. H.

ISAACS, R. B. See LITTLE, K. B.

JOHNSON, D. M. See BRALEY, L. S.

JORDAN, A. E. See SCOTT, T. R.

KANFER, F. H., AND MARSTON, A. R. **Determinants of Self-Reinforcement in Human Learning** 245

KANFER, F. H. See MARSTON, A. R.

KATZ, L. See MYERS, J. L.

KATZ, P. A. **Effects of Labels on Children's Perception and Discrimination Learning** 423

KIRKPATRICK, M. See PETERSON, L. R.

KOPPMAN, J. W., AND GRICE, G. R. **Goal-Box and Alley Similarity in Latent Extinction** 611

KORN, J. H., AND LINDLEY, R. H. **Immediate Memory for Consonants as a Function of Frequency of Occurrence and Frequency of Appearance** 149

LABERGE, D. See GREENO, J. G.

LANE, H. **The Autophonic Scale of Voice Level for Congenitally Deaf Subjects** 328

LEHR, D. J. See BERGUM, B. O.

LEMAY, M. See CHERNIKOFF, R.

LEWIS, J. L. See BERLYNE, D. E.

LEWIS, M., WALL, A. M., AND ARONFREED, J. **Developmental Change in the Relative Values of Social and Nonsocial Reinforcement** 133

LINDLEY, R. H. **Effects of Controlled Coding Cues in Short-Term Memory** 580

LINDLEY, R. H. See KORN, J. H.

LIPSITT, L. P., AND LOLORDO, V. M. **Interactive Effect of Stress and Stimulus Generalization on Children's Oddity Learning** 210

LITTLE, K. B., BRACKBILL, Y., ISAACS, R. B., AND SMELKINSON, N. **A Further Test of a General Utility Theory Model for Probability Learning** 107

LOCKHART, R. A. See GRINGS, W. W.

LOLORDO, V. M. See LIPSITT, L. P.

MARSTON, A. R., AND KANFER, F. H. **Human Reinforcement: Experimenter and Subject Controlled** 91

Marston, A. R. See Kanfer, F. H.
Martin, J. G., Oliver, M., Hom, G., and Heaslet, G. **Repetition and Task in Verbal Mediating-Response Acquisition** 12
Marx, M. H., Tombaugh, J. W., Cole, C., and Dougherty, D. **Persistence of Nonreinforced Responding as a Function of the Direction of a Prior-Ordered Incentive Shift** 542
Minton, H. L. **A Replication of Perceptual Curiosity as a Function of Stimulus Complexity** .. 522
Morin, R. E., and Forrin, B. **Response Equivocation and Reaction Time** 30
Mudd, S. A. **Spatial Stereotypes of Four Dimensions of Pure Tone** 347
Murdock, B. B., Jr. **Interpolated Recall in Short-Term Memory** 525
Musgrave, B. S., Goss, A. E., and Shrader, E. **Compound Nonsense-Syllable Stimuli Presented without an Intervening Space** 609
Myers, J. L., Fort, J. G., Katz, L., and Suydam, M. M. **Differential Monetary Gains and Losses and Event Probability in a Two-Choice Situation** 521
Nodine, C. F. **Stimulus Durations and Stimulus Characteristics in Paired-Associates Learning** 100
Oliver, M. See Martin, J. G.
Passey, G. E., and Wood, D. L. **Effect of Pattern of Reinforcement on the Conditioned Eyelid Response** 241
Peterson, L. R., Wampler, R., Kirkpatrick, M., and Saltzman, D. **Effect of Spacing Presentations on Retention of a Paired Associate over Short Intervals** 206
Peterson, M. J. **Effects of Knowledge of Results on a Verbal Mediating Response** 394
Peterson, M. J., and Blattner, K. C. **Development of a Verbal Mediator** 72
Pierce, J. **Some Sources of Artifact in Studies of the Tachistoscopic Perception of Words** ... 363
Pollio, H. R. **Word Association as a Function of Conditioned Meaning** 454
Powell, D. A. See Scott, T. R.
Pradhan, P. L., and Hoffman, P. J. **Effect of Spacing and Range of Stimuli on Magnitude Estimation Judgments** 533
Pritchard, J. F. See Bevan, W.
Prokasy, W. F., Ebel, H. C., and Thompson, D. D. **Response Shaping at Long Interstimulus Intervals in Classical Eyelid Conditioning** 138
Provins, K. A. See Bell, C. R.
Rainey, C. A. See Bryden, M. P.
Reynolds, W. F., Anderson, J. E., and Besch, N. F. **Secondary Reinforcement Effects as a Function of Method of Testing** 53
Rich, S. See Fleishman, E. A.
Robinson, A. W., and Clayton, K. N. **Effect of Duration of Confinement in a Nonbaited Goal Box on the "Apparent Frustration Effect"** 613
Rosenberg, S. **Behavior in a Continuous-Response Task with Noncontingent Reinforcement** 168
Rutledge, E. F. See Spence, K. W.
Salapatek, P. H. See Berlyne, D. E.
Saltz, E. **Compound Stimuli in Verbal Learning: Cognitive and Sensory Differentiation versus Stimulus Selection** 1
Saltz, E., and Asdourian, D. **Incubation of Anxiety as a Function of Cognitive Differentiation** 17
Saltzman, D. See Peterson, L. R.
Saufley, W. H., Jr., and Bilodeau, I. McD. **Protective Self-Pacing during Learning** 596
Schartz, H. A. **Influence of Instructional Set and Response Frequency on Retroactive Interference** 127
Schiller, P. H., and Wiener, M. **Monoptic and Dichoptic Visual Masking** 386
Scott, T. R., Jordan, A. E., and Powell, D. A. **Does Visual Aftereffect of Motion Add Algebraically to Objective Motion of the Test Stimulus?** 500
Seibel, R. **Discrimination Reaction Time for a 1,023-Alternative Task** 215
Sewall, S. T. See Swets, J. A.
Sgro, F. J. **Beta Motion Thresholds** 281
Sgro, J. A., and Weinstock, S. **Effects of Delay on Subsequent Running under Immediate Reinforcement** 260
Shrader, E. See Musgrave, B. S.
Siegel, S., and Wagner, A. R. **Extended Acquisition Training and Resistance to Extinction** .. 308
Smelkinson, N. See Little, K. B.
Spence, K. W., Rutledge, E. F., and Talbott, J. H. **Effect of Number of Acquisition Trials and the Presence or Absence of the UCS on Extinction of the Eyelid CR** 286
Spivey, J. E. See Capaldi, E. J.
Stevens, S. S., and Guirao, M. **Subjective Scaling of Length and Area and the Matching of Length to Loudness and Brightness** 177
Stone, H. **Determination of Odor Difference Limens for Three Compounds** 466
Suydam, M. M. See Myers, J. L.
Swets, J. A., and Sewall, S. T. **Invariance of Signal Detectability over Stages of Practice and Levels of Motivation** 120
Talbott, J. H. See Spence, K. W.

THOMPSON, D. D. See PROKASY, W. F.
TOMBAUGH, J. W. See MARX, M. H.
TRABASSO, T. See BOWER, G.
TULVING, E., AND GOLD, C. **Stimulus Information and Contextual Information as Determinants of Tachistoscopic Recognition of Words** 319
UHL, C. N. **Learning of Interval Concepts: I. Effects of Differences in Stimulus Weights** 264
UHL, C. N. **Two-Choice Probability Learning in the Rat as a Function of Incentive, Probability of Reinforcement, and Training Procedure** 443
VAN SANT, C. See GOTTSDANKER, R.
WAGNER, A. R. **Conditioned Frustration as a Learned Drive** 142
WAGNER, A. R. See SIEGEL, S.
WAGNER, J. E. See BRACKBILL, Y.
WALL, A. M. See LEWIS, M.
WAMPLER, R. See PETERSON, L. R.
WEINSTOCK, S. See SGRO, J. A.
WEISS, W., AND HODGSON, D. A. **Influence of Unbalanced Response Scales on Size Judgments** 37
WEITZMAN, B. **A Figural Aftereffect Produced by a Phenomenal Dichotomy in a Uniform Contour** 195
WEITZMAN, B. **A Threshold Difference Produced by a Figure-Ground Dichotomy** 201
WICKENS, D. D., ALLEN, C. K., AND HILL, F. A. **Effect of Instructions and UCS Strength on Extinction of the Conditioned GSR** 235
WIENER, M. See SCHILLER, P. H.
WILKINSON, R. T. **Aftereffect of Sleep Deprivation** 439
WILKINSON, R. T. **Interaction of Noise with Knowledge of Results and Sleep Deprivation** ... 332
WINNICK, W. A., AND DORNBUSH, R. L. **Role of Positional Cues in Serial Rote Learning** 419
WOLLEN, K. A. **Relationships between Choice Time and Frequency during Discrimination Training and Generalization Tests** 474
WOOD, D. L. See PASSEY, G. E.
ZEILER, M. D. **New Dimensions of the Intermediate Size Problem: Neither Absolute nor Relational Response** 588

Journal of
Experimental Psychology

Vol. 66, No. 1 July 1963

COMPOUND STIMULI IN VERBAL LEARNING: COGNITIVE AND SENSORY DIFFERENTIATION VERSUS STIMULUS SELECTION[1]

ELI SALTZ

Center for the Study of Cognitive Processes, Wayne State University

Previous studies have shown that paired-associate learning could be facilitated by surrounding each verbal stimulus with a different color. The present studies were designed to determine if such facilitation occurs when the colors cannot readily be used as cues. Critical manipulations involved alternating learning trials (S-R pairs presented simultaneously) and test trials (only stimuli presented). Availability of colors as cues was reduced by presenting them during learning trials but not test trials, or during test trials but not learning trials. 310 Ss in various experimental conditions indicated color differentiation can significantly facilitate learning even when color cannot readily be used as a cue.

Weiss and Margolius (1954), in a study of context cues in verbal paired-associate learning, found that learning was facilitated if each stimulus was printed on a different colored background. On the basis of the writer's theoretical position (Saltz, 1961; Saltz & Riach, 1961), these results can be interpreted in terms of stimulus differentiation. However, this interpretation is clouded by the fact that *S*s in the Weiss and Margolius study could use the colored backgrounds per se as cues to the appropriate responses. The results can be interpreted as indicating that color was a more facilitating cue than nonsense syllables. Indeed, on a retention test 24 hr. after learning, Weiss and Margolius found that recall to colors alone was significantly better than recall to nonsense syllables alone, though colors plus nonsense syllables produced the greatest recall. Underwood (1963) has interpreted the Weiss and Margolius results in terms of cue selection—*S*s select aspects of the stimulus compound as the functional cue. Underwood's data are consistent with those of Weiss and Margolius in finding that both color alone and nonsense syllable alone will evoke correct responses following original learning in which syllables were surrounded by colored backgrounds.

While the studies cited above suggest that both color and syllable aspects of a compound stimulus may acquire cue properties, they do not unambiguously show that this accounts for the superior performance of *S*s who have compound cues available. A more direct test of differentiation vs. stimulus selection would involve testing the effects of context stimuli when their cue properties are removed. This was accomplished in the present study by employing separate learning and test trials. On learning trials, the stimulus and response terms were presented simultaneously. On test trials, only the stimuli were presented. Thus stimuli could be on colored backgrounds during learning trials but not during test trials; in this condition, color could become associated with the correct response, but the color cue would not be available during test. Or, color could be present on the test trials but not on the learning trials; in this condition, color could not readily become associated with the correct responses.

---

[1] The present study is part of a program of research, concerned with the influence of differentiation on verbal processes, supported by National Science Foundation Grant G21476.

This type of experimental manipulation permits analysis of the locus of a differentiation effect. Should learning be facilitated for the condition in which color is absent during learning but is present during test, differentiation could be assumed to be a perceptual variable produced by sensory distinctiveness between stimuli. This would be similar to the typical discrimination learning situation in which learning becomes easier as discriminanda become less similar. On the other hand, should learning be facilitated for the condition in which color is present during learning but not present during test trials, differentiation could not be assumed to be a property of the objective stimuli present on the test trials. In this case some type of distinctiveness could be assumed to have been acquired during the learning trials and retained as a learned factor during the test trials.

Several alternative mechanisms could be suggested for such an acquired distinctiveness variable (henceforth to be referred to as cognitive differentiation in the present paper). However, at this time the critical consideration is that any obtained facilitation of paired-associate learning in this condition could not be readily accounted for by transfer of the association between color and relevant verbal list response from learning trials to test trials. Instead, it would appear to indicate a learned alteration in *S*'s internal representation of the relevant stimuli, an alteration which increases the distinctiveness of the stimuli.

EXPERIMENT I

Method

General design.—A basic 2 × 2 design was employed which involved color context cues present or absent during learning trials, and present or absent during test trials. This design was repeated in each of three different replications. In Replication 1, the context stimuli were squares of color approximately the area of the verbal stimuli, and pasted on the memory drum tape immediately to the right of the verbal stimuli. In Replications 2 and 3, a rectangular border of color approximately $\frac{1}{16}$ in. thick was drawn around each appropriate stimulus word. For a given *S*, the same color accompanied any given verbal stimulus on all trials. Replications 1 and 2 employed the same verbal S-R pairs, while Replication 3 employed a different set of verbal materials. All *S*s were given 15 learning trials and the number of correct responses over these trials was analyzed in a 2 × 2 × 3 analysis of variance (i.e., Color during Learning Trials × Color during Test Trials × Replications).

Subjects.—The *S*s were 160 introductory psychology students at Wayne State University. Replication 1 employed 80 *S*s; Replications 2 and 3 employed 40 *S*s each. The *S*s were randomly assigned to conditions with the limitation that in each experiment the number of *S*s in each condition was equal.

TABLE 1

MEAN NUMBER CORRECT RESPONSES OVER 15 TEST TRIALS FOR EACH OF THREE REPLICATIONS: EXP. I

Cond.		Replication 1[a]		Replication 2[b]		Replication 3[b]		Grand M
Learning	Test	M	SD	M	SD	M	SD	
Color	Color	86.4	6.12	101.9	6.43	102.7	7.20	94.4
Color	No Color	67.6	6.30	99.1	10.10	76.9	11.72	77.8
No Color	Color	62.4	6.22	79.1	9.59	91.6	5.86	73.8
No Color	No Color	59.7	7.08	50.8	8.51	69.9	5.30	60.1

[a] $N = 20$ per condition.
[b] $N = 10$ per condition.

Materials.—Each list consisted of 10 paired associates. The pairs in Replications 1 and 2 were QOR-MODERN, BEK-TIRESOME, HAQ-SPICY, DUF-PLASTIC, SOZ-FAITHFUL, KIV-WOUNDED, WIH-FIERY, FEJ-GUILTY, VAW-HAIRY, JUY-EQUAL. The pairs in Replication 3 were FAH-STERILE, GIZ-NEUTER, CES-HYBRID, NOL-INFORM, YAD-VULGAR, MIJ-CRUMBLING, QUN-WARLIKE, KOR-PLEASANT, VUT-WORLDLY, WEH-FLASHY. Pairs were presented in three different random orders in each experiment. Colors employed were red, royal blue, green, gray, black, tan, brown, yellow, violet, and powder blue.

Procedure.—The *S*s were instructed to learn the word that accompanied each nonsense syllable so that when the nonsense syllable appeared alone, they could indicate the appropriate word. No reference was made concerning color during the experiment. Each of the 15 learning trials was followed by a test trial. During learning trials, each S-R pair was presented simultaneously for 2 sec. During test trials, only the stimulus member of each pair was presented, presentation rate of 2 sec. per item. A 4-sec. interval was employed between alternate learning and test trials.

Results

Table 1 summarizes the mean number of correct responses in 15 trials for each condition in each experiment. In each of the three replications, Color-Color is the best condition, No Color-No Color the worst, with the Color-No Color and the No Color-Color intermediate. The analysis of variance over these data is summarized in Table 2. Both cognitive differentiation (i.e., color during learning trials) and sensory differentiation (i.e., color during test trials) significantly facilitate learning. Since in the Color-Color condition, the various colors could serve as cues as well as have differentiation properties, the critical consideration is whether the Color-No Color condition and the No Color-Color condition facilitate learning. That they do facilitate learning is indicated by the fact that the CL × CT interaction does not approach significance. If the significant main effects due to color during learning or color during test trials were produced solely by the Color-Color condition or by the Color-Color plus only one of the other experi-

TABLE 2

ANALYSIS OF VARIANCE FOR MEAN CORRECT RESPONSES OVER 15 TRIALS FOR THREE REPLICATIONS: EXP. I

Source	df	F
Color, Learning (CL)	1	19.14***
Color, Test (CT)	1	12.02**
Replications (R)	2	5.94**
CL × R	2	2.55
CL × CT	1	.095
CT × R	2	.74
CL × CT × R	2	1.89
Within cells (error MS)	148	(766.55)

** $p = .01$.
*** $p = .001$.

mental conditions (i.e., either Color-No Color or No Color-Color) the CL × CT interaction would be significant. As additional verification of the critical relationships, simple effects were calculated for Color-No Color vs. No Color-No Color, $F(1, 74) = 8.27$, and for No Color-Color vs. No Color-No Color, $F(1, 74) = 4.99$. Both effects were significant at well beyond the .05 level. Thus, both cognitive and sensory differentiation proved effective even when the differentiated stimuli could not function as cues.

The nonsignificant CL × R and CT × R interactions indicate that the effect of color differentiation is general across the types of materials used in the present investigation. The nonsignificant triple interaction (CL × CT × R) indicates that the relative shifts in position of the Color-No Color and the No Color-Color conditions between replications is no greater than one would expect by chance if these two conditions produced effects of equal magnitude.

EXPERIMENT II

To accommodate the results of Exp. I, a cue selection theory could reason: In the Color-No Color condition, *S*s might associate the color with both the CVC stimulus and the adjective response during the learning trials; on the test trials the CVC could then evoke the color which could mediate the adjective. In the No Color-Color condition, *S* might associate the color with the CVC stimulus during the test trials; on the learning trials, the CVC could then evoke the color which could become associated with the adjective response; on subsequent test trials the color could serve as a cue to the correct response. The writer believes that such activity on the part of *S* is too complex to be reasonably assumed. However, Exp. II was designed to reduce the effectiveness of such a strategy. To accomplish this, verbal stimuli were selected which were easier to associate to the response items than were the colors. Consequently, development by *S* of a strategy which would permit utilization of the colors as stimuli in place of the verbal stimuli would not facilitate learning.

TABLE 3

MEAN NUMBER CORRECT RESPONSES OVER 15 TEST TRIALS: EXP. II

Test	Learning				
	Color		No Color		Grand *M*
	M	*SD*	*M*	*SD*	
Color	99.57	25.3	92.30	25.5	95.94
No Color	90.20	23.4	76.83	24.7	83.51
Grand *M*	94.88		84.56		

Method

Subjects.—The *S*s were 150 students in introductory psychology at Wayne State University, 30 *S*s assigned randomly to each of five groups.

Materials.—Four of the five lists in Exp. II consisted of the following 10 paired associates: FUR-VAW, NUT-SOZ, BUS-JUY, SEX-KIV, LAW-FEJ, DAY-HAQ, TOM-DUF, JOB-WIH, MAD-BEK, WIN-QOR. The stimuli were selected from Noble's (1961) revised CVC norms. A pilot study indicated these pairs to be more rapidly learned than pairs consisting of the colors in Exp. I as stimuli associated to the above responses. As in Exp. I, color was employed with this basic list to produce four different lists: color plus CVC as stimulus on both learning and test trials; color plus CVC on learning trials, CVC alone on test trials; CVC alone on learning trials, color plus CVC on test trials; and no color during either learning or test trials. The fifth list consisted of color alone as stimulus on both learning and test trials. The colors were those used in Exp. I, and their application was that of Replications 2 and 3.

Procedure.—The procedure was identical with that of Exp. I.

Results

The results, summarized in Table 3, parallel very closely those obtained in Exp. I. Color-Color produces the best learning, No Color-No Color the worst, while Color-No Color and No Color-Color are intermediate and approximately equal in effect. Main effects due to color during learning are significant at beyond the .01 level, $F(1, 116) = 7.34$, as are main effects due to color during test trials, $F(1, 116) = 10.13$, while the L × T interaction is far from significant, $F(1, 116) = 0.98$. Individual *t*'s indicate that both Color-No Color and No Color-Color conditions exceed the No Color-No Color condition at well beyond the .05 level ($t = 2.13$ and 2.46, respectively). Finally, the pilot data indicating that color alone as a stimulus produced poorer learning than word alone are sustained. The mean correct for color alone was 63.93, that for word alone was 76.83, a difference significant at beyond the .05 level ($t = 2.05$).

Approximately 20 *S*s in the No Color-Color and the Color-No Color conditions were interviewed after the learning trials, concerning their use of color. Several *S*s stated that they attempted to use color as a cue during the first few trials, but all found this too confusing and ceased.

DISCUSSION

The writer believes that the principal contribution of the present study is that it operationally distinguishes between the effects of cognitive and sensory differentiation and indicates that both may facilitate verbal learning. Both the learning data of Exp. I and II and the interview data in Exp. II are strongly opposed to a cue selection interpretation of either sensory or cognitive differentiation.

In the present study, the greatest amount of facilitation occurred in the Color-Color condition. Since this is the only condition in which color could readily act as a cue (viz., color could become associated to the relevant response during learning, and was also available to serve as a cue during test trials) it might be concluded that stimulus selection as well as differentiation acted to facilitate learning. On the other hand, the Color-Color condition is the only one in which both cognitive and sensory differentiation operate simultaneously. The possibility remains, therefore, that the superiority of the Color-Color condition may be accounted for by assuming a simple additive relationship between cognitive and sensory differentiation. Thus, while the Weiss and Margolius (1954) data are clear in indicating that both color and verbal stimuli may acquire cue properties in a Color-Color condition, it is by no means necessary that this double cue property is the one leading to the superiority of the Color-Color *S*s. The data of the present studies are not critical for testing assumption of additivity.

REFERENCES

NOBLE, C. E. Measurements of association value (a), rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. *Psychol. Rep.*, 1961, **8**(Monogr. Suppl. No. 3), 487–521.

SALTZ, E. Response pretraining: Differentiation or availability? *J. exp. Psychol.*, 1961, **62**, 583–587.

SALTZ, E., & RIACH, W. D. The effect of stress on stimulus differentiation. *J. exp. Psychol.*, 1961, **62**, 588–593.

UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. Cofer (Ed.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963.

WEISS, W., & MARGOLIUS, G. The effect of context stimuli on learning and retention. *J. exp. Psychol.*, 1954, **48**, 318–322.

(Received August 22, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 6–11

ROLE OF KINESTHETIC AND SPATIAL-VISUAL ABILITIES IN PERCEPTUAL-MOTOR LEARNING[1]

EDWIN A. FLEISHMAN AND SIMON RICH

Yale University

Ss were administered a spatial test and a new measure of "kinesthetic sensitivity," and then received extended practice on a Two-Hand Coordination (THC) task. The results confirm the hypothesis that sensitivity to proprioceptive cues are more important later in perceptual-motor learning while sensitivity to exteroceptive (spatial-visual) are more critical earlier in learning. The study extends previous work which showed that abilities which contribute to learning early in practice may be different from those which facilitate later learning.

It is generally recognized that responses to kinesthetic stimuli play a role in perceptual-motor skill performance. But relatively little is known about the manner in which these cues operate or about their relative importance during the course of motor skill learning. The main difficulty, of course, is that the relevant proprioceptive stimuli are inaccessible to direct experimental manipulation.

Attempts to investigate such cues, at least remotely, generally manipulate physical features of a control in terms of such variables as mass, spring loading, friction, viscous damping, and amplitude of control displacement (e.g., Bahrick, 1957). Especially illuminating are the recent studies by Adams and associates (e.g., Adams & Creamer, 1962) which attempt to specify, more precisely, the role of proprioception in skilled motor behavior. Their work indicates that proprioception serves not only a regulatory function (feedback for overt responses) but also as a time-perception mechanism which assists *S* in anticipating temporal regularities in stimulus events. Still unanswered are questions of the relative contribution of kinesthetic and nonkinesthetic factors to the learning of a complex perceptual-motor skill, the learning of which may be dependent on responses to *many* classes of stimuli.

The fact that kinesthetic factors play a role in motor performance raises questions of individual differences in ability to make use of proprioceptive cues, and the interaction of such differences with motor skill learning. Are there stable individual differences in sensitivity to kinesthetic cues? Jenkins (1951) stresses that "kinesthesis—the sense of position and movement, is probably the most important sensitivity that man possesses." Yet, we know almost nothing about the existence of possible "kinesthetic ability" traits which may distinguish one individual from another.

A line of research by the author and associates (summarized most recently in Fleishman, 1962) has shown that the particular combinations of abilities contributing to performance on complex tasks may change as practice on these tasks continues. The approach used in these studies is to exploit the fact of individual differences to gain insights into the mechanisms of skill learning.[2] The *S*s' pretask abilities become major "treatment" variables with significant interactions with learning trials. The magnitude of the differences produced by these "organismic" trait variables are often quite large (e.g. Fleishman & Hempel, 1955) when compared with the small effects often achieved by the manipulation of procedural or environmental variables.

---

[1] This study was carried out with the support of National Science Foundation Grant G-23840. The article was written during the senior author's year as a Guggenheim Fellow. This author is indebted to the John Simon Guggenheim Foundation as well as to Yale University, which provided a Senior Faculty Fellowship during this year.

Lately, this program has turned toward the search for external reference measures which might better describe the variance at advanced levels of learning. One of our hypotheses was that at least part of this variance is "kinesthetic" in nature (Fleishman, 1956, 1957a, 1959, 1962) and represents individual differences in sensitivity to kinesthetic cues.

Fitts (1951), stressed the need to investigate the relative importance of exteroceptive vs. interoceptive feedback, and the optimum combination of both, in motor performance. He goes on to say that "visual control is important while an individual is learning a new perceptual-motor task. As performance becomes habitual, however, it is likely that proprioceptive feedback or 'feel' becomes the more important." Fitts' hypothesis is directly in line with our own speculation growing out of our extended practice correlational studies. But as far as we know this hypothesis has never been subjected to experimental test.

The present study represents a preliminary attempt to investigate this hypothesis within the framework of our ability-skill paradigm. According to our reasoning, if kinesthetic cues predominate, mainly, later in motor learning, then *S*s who have superior sensitivity to these cues should be superior to other *S*s at advanced stages of learning a complex motor task; but these *S*s would not necessarily excel during initial stages of learning. Also tested was the related hypothesis that early in learning, spatial-visual abilities would play a dominant role in perceptual-motor learning, but that these abilities would decrease in importance as practice continued.

Reliable measures of spatial abilities exist but first, we needed to develop a reliable measure of individual differences in "kinesthetic sensitivity."

METHOD

The *S*s were 40 males, all Yale University undergraduates enrolled in a second semester psychology course. Each *S* received extended practice on the Two-Hand Coordination apparatus and in addition was administered two ability measures: a standardized test of spatial orientation and a newly developed test of "kinesthetic sensitivity."

Ability Measures

Kinesthetic sensitivity measure.—The measure used was based on the classical psychophysical procedure of determining difference limens for judgments of lifted weights. Such a measure is relatively easy to administer and brings into play some of the same hand and arm muscles to be used with the Two-Hand Coordination (THC) apparatus. Despite the widespread use of these materials in classical studies of difference thresholds, little is known about the reliability of such measures as indicators of individual differences. Variability among individuals in threshold determinations, is error variance in psychophysical study, but it is precisely those differences that we are interested in here. We were not even sure that kinesthetic sensitivity, as measured in this way, represents a dependable ability trait. Considerable pretesting was carried out to arrive at the desirable number of trials, the range of weights to be used, the specific instructions to give, and the type of score to derive.

---

[2] The ability-skill paradigm used here has been elaborated elsewhere (Fleishman, 1956, 1959, 1962; Gagné & Fleishman, 1959). Briefly, the term *ability*, as used here, refers to a more general, stable trait of the individual which may facilitate (transfer to) performance in a wide variety of different tasks. The term *skill* is more specific, referring to level of proficiency on a specific task. Thus, flying an airplane is a skill, while spatial orientation is a more general ability related to performance on many diverse tasks (e.g. dentistry, navigation). Of course, abilities themselves are often products of earlier learning.

The weights used were identical (in size) brass cylinders of 100, 102, 104, 106, 108, 110, and 112 gm. The standard was the 106-gm. weight. The method of limits was used, in which blindfolded *S*s compared each variable weight in a pair with the standard, making the judgment of whether the second weight was lighter, heavier, or the same, as the first in each pair. Each *S* was blindfolded and received 24 such pairs, where each weight was lifted one at a time. The *S*'s arm was on the table with lifting done from the wrist. The difference limen (DL), calculated for each *S* was the measure of kinesthetic sensitivity used. The procedure for deriving this is outlined in Woodworth and Schlosberg (1954, pp. 197–198).

Test-retest reliability of the DLs derived by this test procedure, with at least 24 hr. between the two administrations was, .85. As far as we know, this is the first demonstration that individuals vary consistently in their DLs.

Spatial ability measure.—The measure used was the United States Air Force Aerial Orientation Test found to load consistently on a spatial orientation factor in many previous factor analyses (Fleishman, 1957b; Guilford & Lacey, 1947; Michael, Guilford, Fruchter, & Zimmerman, 1957; Parker & Fleishman, 1960; Roff, 1951). The test presents a printed booklet containing a series of aircraft cockpit views of land-sea-sky horizons. For each item *S* is required to match the cockpit view of the horizon that corresponds to the position of one of five airplanes from which this view would be seen. Score is the number right. The reliability of this test, in these previous studies, averages approximately .84.

Perceptual-Motor Practice Task

The Two-Hand Coordination apparatus has been described in detail by Melton (1947), and is pictured in reports by Fleishman (1956, 1958, 1960), Parker and Fleishman (1960), and in Gagné and Fleishman (1959). Briefly, *S* attempts to keep a target-follower on a small target disc as the target moves irregularly and at various rates around a circular plate. Movement of the target-follower to the right and left is controlled by one lathe-type handle; to-and-from movement of the target-follower is controlled by the other handle. Consequently, simultaneous rotation of both handles moves the follower in any *resultant* direction. Score is the total time-on-target during a trial.

In the present study *S*s received 40 1-min. trials. These were separated by 15-sec. rests, except after Trials 8, 16, and 32 when there was a 1-min. rest.

RESULTS

Prediction of Two-Hand Coordination performance from the ability measures.—The correlations between the Aerial Orientation Test (number correct) and kinesthetic sensitivity (DL) measures, with total time-on-target during the 40 trials of Two-Hand Coordination performance were found to be .49 and .58, respectively. Both of these coefficients are significant at the .01 level. The correlation between Aerial Orientation and the DL was .12, not statistically significant. These results indicate that our two ability measures, tap independent ability traits, each of which accounts for substantial variance in the perceptual-motor practice task. The combination of these two tests yields a multiple *R* of .73 with overall Two-Hand Coordination performance. Thus, over half the variance ($R^2 = .53$) in total time-on-target on the perceptual-motor task, is "accounted for" in

TABLE 1

CORRELATIONS OF SPATIAL AND KINESTHETIC TESTS WITH SUCCESSIVE TRIALS OF TWO-HAND COORDINATION TASK PERFORMANCE

THC Trial	Aerial Orientation	Kinesthetic Sensitivity
1	.36**	.03
2	.28*	.19
3	.22*	.15
4	.19	.15
5	.08	.10
6	.07	.09
7	.09	.23*
8	−.05	.28*
9	−.02	.38**
10	.01	.40**

* $p < .05$, one-tailed.
** $p < .01$, one-tailed.

terms of two ability traits developed by *S* prior to practice with this task.

Relation of ability measures to different stages of THC learning.—The 40 THC trial scores were grouped into 10 blocks of four trials each. Correla-

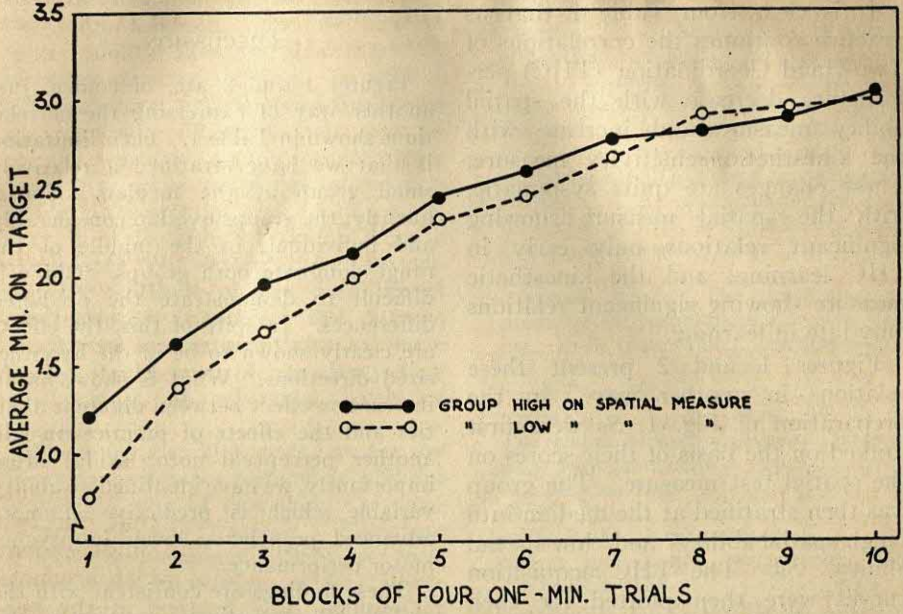


FIG. 1. Comparison of Two-Hand Coordination acquisition curves for groups high and low on the Aerial Orientation Test.

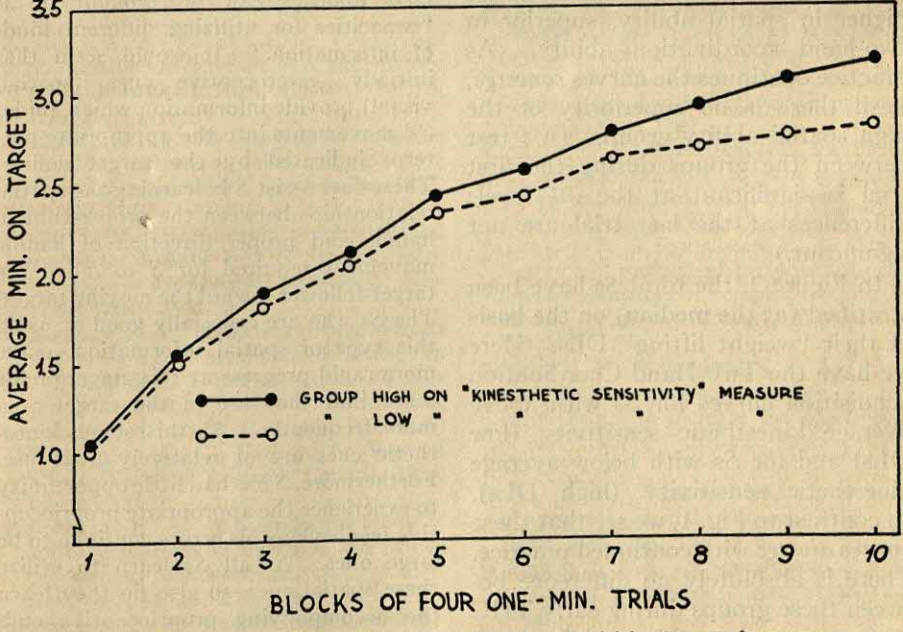


FIG. 2. Comparison of Two-Hand Coordination acquisition curves for groups high and low on the Kinesthetic Sensitivity measure.

tions of each of these 10 time-on-target scores with the ability tests were obtained. These are presented in Table 1.

It is clear from Table 1 that as practice continues the correlations of Two-Hand Coordination (THC) performance decrease with the spatial ability measure and increase with the kinesthetic sensitivity measure. These changes are quite systematic, with the spatial measure showing significant relations only early in THC learning, and the kinesthetic measure showing significant relations only late in learning.

Figures 1 and 2 present these relations in another way. In the preparation of Fig. 1, *S*s were first ranked on the basis of their scores on the spatial test measure. The group was then stratified at the median into "high spatial ability" and "low spatial ability" *S*s. The THC acquisition curves were then plotted for each group. Figure 1 shows that the groups are clearly separated at the beginning of learning, with those *S*s higher in spatial ability, superior in two-hand coordination ability. As practice continues the curves *converge*, until there is no superiority of the high spatial ability group. (A *t* test between the groups during the first trial is significant at the .01 level; differences at the last trial are not significant.)

In Figure 2, the total *S*s have been stratified (at the median) on the basis of their "weight lifting" DLs. Here we have the Two-Hand Coordination acquisition curves for *S*s with above average kinesthetic sensitivity (low DLs) and for *S*s with below average kinesthetic sensitivity (high DLs). In contrast to Fig. 1, we see that these curves *diverge* with continued practice. There is absolutely no difference between these groups during early THC performance, but this variable produces a significant difference later in learning. (A *t* test between the groups at the last trial is significant at the .01 level.)

DISCUSSION

Figures 1 and 2 are, of course, just another way of expressing the correlations shown in Table 1. Their limitation is that we have stratified a relatively small group at the median. Consequently, the groups overlap considerably and individuals in the middle of the range dominate both groups, making it difficult to demonstrate the predicted differences. In spite of this, the effects are clearly shown to be in the hypothesized directions. What is shown is an interaction effect between different abilities and the effects of practice on still another perceptual-motor skill. More importantly, we have identified an ability variable which is predictive of more advanced proficiency levels in psychomotor performance.

These findings are consistent with the ability-skill paradigm, previously described, as well as with an information processing model of human learning. In fact, abilities can be thought of as "capacities for utilizing different kinds of information." It would seem that initially, exteroceptive cues (spatial-visual) provide information which guides *S*s' movements into the appropriate patterns indicated by the target course. These cues assist *S* in learning the spatial relationships between the proper control handle and proper direction of handle movement required for *S* to move the target-follower toward the moving target. The *S*s who are especially good at using this type of spatial information, make more rapid progress at this stage, in the sense that they are in the target area more frequently. At this stage, kinesthetic cues are of relatively little use. Furthermore, *S* has had little opportunity to experience the appropriate proprioceptive feedback as his errors continue to be large ones. As all *S*s learn to utilize these spatial cues, so also do they learn the accompanying proprioceptive cues which accompany the movements originally controlled by the exteroceptive feedback. Once a given level of proficiency is reached and errors tend to be smaller, spatial cues are not as effective in facilitating more precise control over the target-follower. The *S*s high in spatial ability have an advantage only in earlier stages of learning. To achieve a higher score, finer motor adjustments are required. Those *S*s who are especially sensitive to proprioceptive cues are able to make use of this information earlier in the practice period. In a sense, they may be able to switch from a dependence on exteroceptive (visual-spatial) feedback to the more direct proprioceptive channels. The *S*s who are limited in this sensitivity are more limited in the level of proficiency they can achieve at advanced levels of practice.

The study also illustrates how the use of external reference tests of *S*s' abilities provides a way of describing the components of skill learning. Elsewhere (Parker & Fleishman, 1961) we have shown that knowledge of how these components interact with practice at different stages of skill learning can facilitate skill training.

REFERENCES

Adams, J. A., & Creamer, L. R. Proprioception variables as determiners of anticipatory timing behavior. *Hum. Factors*, 1962, **3**, 217–222.

Bahrick, H. P. An analysis of stimulus variables influencing the proprioceptive control of movements. *Psychol. Rev.*, 1957, **64**, 324–328.

Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Fleishman, E. A. Predicting advanced levels of proficiency in psycho-motor skills. In G. Finch & F. Cameron (Eds.), *Air force human engineering, personnel, and training research*. (NAS-NRC Publ. No. 455) Washington, D. C.: National Academy of Science-National Research Council, 1956. Pp. 142–151.

Fleishman, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. *J. appl. Psychol.*, 1957, **41**, 263–272. (a)

Fleishman, E. A. Factor structure in relation to task difficulty in psychomotor performance. *Educ. psychol. Measmt.*, 1957, **17**, 522–532. (b)

Fleishman, E. A. Dimensional analysis of movement reactions. *J. exp. Psychol.*, 1958, **55**, 483–453.

Fleishman, E. A. Abilities and the learning of psychomotor skills. In P. H. Dubois, W. H. Manning, & C. J. Spies (Eds.), Factor analysis and related techniques in the study of learning. Technical Report No. 7, 1959, Washington University, Contract Nonr-816(02), Office of Naval Research.

Fleishman, E. A. Psychomotor tests drug research. In J. G. Miller & L. Uhr (Eds.), *Drugs and behavior*. New York: Wiley, 1960.

Fleishman, E. A. The description and prediction of perceptual-motor skill learning. In R. Glaser (Ed.), *Training research and education*. Pittsburgh: Univer. Pittsburgh Press, 1962.

Fleishman, E. A., & Hempel, W. E., Jr. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, **49**, 301–310.

Gagné, R. M., & Fleishman, E. A. *Psychology and human performance*. New York: Holt, 1959.

Guilford, J. P., & Lacey, J. T. (Eds.) *Printed classification tests*. Washington: United States Government Printing Office, 1947.

Jenkins, W. O. Somesthesis. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Melton, A. W. (Ed.) *Apparatus tests*. Washington: United States Government Printing Office, 1947.

Michael, W. B., Guilford, J. P., Fruchter, B., & Zimmerman, W. S. The description of spatial-visualization abilities. *Educ. psychol. Measmt.*, 1957, **17**, 185–199.

Parker, J. F., & Fleishman, E. A. Ability factors and component performance measures as predictors of complex tracking behavior. *Psychol. Monogr.*, 1960, **74**(16, Whole No. 503).

Parker, J. F., & Fleishman, E. A. Use of analytical information concerning task requirements to increase the effectiveness of skill training. *J. appl. Psychol.*, 1961, **45**, 295–302.

Roff, M. F. Personnel and classification procedures: Spatial tests. *USAF Sch. Aviat. Med. Rep.*, 1951, No. 21-29-002.

Woodworth, R. L., & Schlosberg, H. *Experimental psychology*. New York: Holt, 1954.

(Early publication received February 18, 1963)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 12–16

REPETITION AND TASK IN VERBAL MEDIATING-RESPONSE ACQUISITION

JAMES G. MARTIN, MICHAEL OLIVER, GEORGE HOM, AND GARY HEASLET

Chico State College

The 4-list A-B, C-B, A-D, C-D paradigm for mediated learning was studied. The first 2 lists were assumed from word association norms. A and C words on the last 2 lists were paired with a single motor response in recognition learning. Starting from the hypothesis that A-D list learning was an interference situation, practice levels and tasks were varied during the A-D list with 64 Ss. Important results on the last list were: (a) mediated words were learned faster than controls, and (b) the mediation effect diminished with increased practice on the mediator-acquisition (A-D) list.

Russell and Storms (1955) have made a clear case for learning mediated by word chains derived from word association norms. Taking words from norms linked in B-C-D fashion, they required *S*s to learn B as a response to a nonsense syllable stimulus, A. Then *S*s learned A-D pairs. Although no *S* reported awareness of any of the inferred associations, the A-D pairs were learned faster than A-X controls. The mediators, it is assumed here, were implicit responses with stimulus properties, that is, cue-producing responses.

Jenkins (1963) has described other verbal mediating connections besides word chains which could be investigated. One is the relation between stimulus words having a common response word which may act as a mediator. This word relation differs substantially, however, from the word chains of Russell and Storms. The steps in the Russell-Storms experiment may be conceived this way: (*a*) Infer B-C and C-D; (*b*) learn A-B; (*c*) learn A-B-C-D with the help of the mediating links B-C-D. When two words having an inferred common response are studied the steps can be conceived: (*a*) A-B and C-B are inferred; (*b*) A-D is learned overtly, but B-D is learned also when B occurs implicitly; (*c*) C-B-D is facilitated via the mediator B. The latter paradigm is like the one used in the investigation by Shipley (1935), upon which Hull (1939) based his paper on mediated stimulus equivalence. One difference, perhaps important, between the Russell-Storms paradigm and the Shipley paradigm is that in Step *b* of the latter the potential mediator B is presumed to occur while the overt response D is being learned, a situation in which, however, B may actually act as a response interfering with D (Jenkins, 1963). The first question in this study is whether the B term can be shown to be an effective mediator if it must acquire mediating properties under interference conditions.

The second question concerns the effects of trials and tasks on acquisition of a mediating response. Given that B-D can form under interference conditions, will the number of trials during which B occurs affect its mediating potentiality? Barnes and Underwood (1959), using the interference paradigm A-B, A-C, found decreasing recall of A-B while A-C was

TABLE 1

WORDS USED IN THE EXPERIMENT

A Words	C Words	B Words	F_L Words	F_T Words	E Words
THIRSTY[a] (34.5)	OCEAN (31.1)	WATER	CHEESE	ANGER	CABBAGE
BLOSSOM[a] (66.7)	STEM (39.9)	FLOWER	SLEEP	EARTH	EAGLE
HARD[a] (66.9)	LOUD (53.7)	SOFT	CITY	HEALTH	JOY
TABLE[a] (83.3)	ROCKING[b] (56.3)	CHAIR	THIEF	JUSTICE	BITTER
BLACK[a] (74.5)	SNOW[b] (25.2)	WHITE	MUTTON	NEEDLE	KING
DARK (82.2)	LAMP (62.8)	LIGHT	BIBLE	WISH	
LONG (75.2)	TALL[b] (50.4)	SHORT	STOMACH	FOOT	
SLOW (74.6)	RUN[b] (30.2)	FAST	BOY	CHILD	
SCISSORS (66.6)	KNIFE[b] (30.2)	CUT	CITIZEN	FRUIT	
TOBACCO (51.1)	CHIMNEY[b] (21.8)	SMOKE	COMMAND	LION	

Note.—Numbers in parentheses, expressed as percentages, are associative frequencies with the B word in the same row.

[a] A words on the first form of the learning list.

[b] Local norms, $N = 119$.

learned and concluded that the A-B association was progressively unlearned. Unlearning during interference conditions seems to apply even though the unlearned associations are extraexperimental (Postman, 1961). Task variables also may assert themselves in mediating-response acquisition. An earlier, related study of mediated generalization in the A-B, A-C, B-C paradigm (Martin, 1961) showed that when the A-C list task was discrimination learning, a situation requiring mediator acquisition under interference conditions, less test-list generalization resulted than after a simpler task in which the A-C list was presented repeatedly for memorization.

The focus in this report is upon the conditions under which mediating associations are formed. From considerations above it was predicted that mediated learning of the last list in the A-B, C-B, A-D, C-D paradigm (*a*) would occur, but (*b*) in inverse relation to practice with the A-D list and (*c*) less when the A-D list task was discrimination learning.

METHOD

Subjects and apparatus.—Sixty-four extra-credit volunteers from introductory courses were assigned randomly to four conditions and, within conditions, to one of two word lists. Most had been in a previous verbal learning experiment in which, however, different verbal materials were used. A memory drum set behind a 2½ × 3 ft. black screen presented stimulus words at 4-sec. intervals. A motor response lever similar to one used by Mink (1957) was made of a 6-in. steel rod and protruded 3 in. from a small box mounted in front of the screen. The lever, pushed to the right 10° or more, activated a small light in front of *E*.

Words.—Ten A-B, C-B word-pair combinations from Minnesota norms for the Kent-Rosanoff Word Association Test (Russell & Jenkins, 1954) and from local norms were chosen such that A and C as stimulus words elicit B as a response word but A and C elicit each other only negligibly (less than 1%, except for one A-C pair, SCISSORS-KNIFE, with associative frequency 3.3%) (see Table 1). For example, THIRSTY elicits WATER and OCEAN elicits WATER, but OCEAN and THIRSTY do not elicit each other. Five words designated extras (E) were chosen also from the Minnesota norms, plus 10 learning-list filler words (F_L) and 10 test-list fillers (F_T). No E or F word was associated with any other or with any A, B, or C word with a frequency as great as 3% in either direction. The median frequency was much less than 1%.

Design and procedure.—The first two lists of the four-list paradigm A-B, C-B, A-D, C-D were norm-inferred. The third list, hereafter designated the learning list, consisted of stimulus words to be associated with a single motor response (D). The learning list was presented under four different condi-

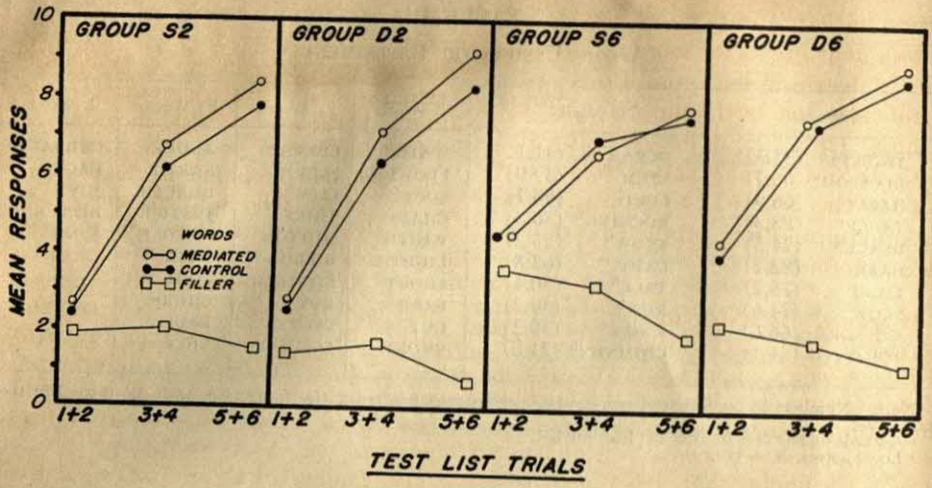


FIG. 1. Effects of learning-list conditions on test-list responses.

tions in a 2 × 2 design using two tasks and two practice levels.

For the simple-learning task, five A words and the five E words were used on the list. The list was continuous and words were randomized within each trial. Instructions to *S* were, in effect, to (*a*) say the words aloud, (*b*) push the lever for each, and (*c*) memorize each word for later recognition.

For the discrimination-learning task, the list consisted of 5 A words, the 5 E words, and the 10 F_L words randomized differently within trials presented continuously. Instructions to *S* were, in effect, (*a*) to say the words aloud and (*b*) to learn (by the anticipation method) to push the lever only for "right" words; i.e., words followed 2 sec. later by a red dot revealed when the memory-drum shutter fell for an additional 2 sec. The red-dot words were the A and E words. No red dots followed the fillers (F_L).

Half the *S*s in each task condition received two trials on the learning list, the other half, six trials, giving a two-trial simple learning group (S2), a six-trial simple (S6), a two-trial discrimination (D2), and a six-trial discrimination group (D6). Two versions of the learning list were constructed. One list using the first five A words of Table 1 was used for half the *S*s in each group. The other five A words were used for the other half of the *S*s.

The fourth list (C-D), hereafter designated the test list, was exactly the same for all *S*s and consisted of the 10 C words (red-dot words) and the 10 F_T words (no red dot). Instructions to *S* were nearly identical to those used in the discrimination-learning condition for the learning list. Since only half the A words appeared on the learning list for any given *S*, only half the C words on the test list *functioned* as mediation words. The other half functioned as controls, with the functions reversed for half the *S*s, providing counterbalancing for irrelevant stimulus properties of the words. The test list was presented continuously for six trials, each randomized differently. Test responses for a given *S* were scored in terms of word function.

RESULTS AND DISCUSSION

Figure 1 shows mean response curves for each type of word on the test list. Of particular interest, of course, are the differences between mediated and control words.

Analysis of variance of total responses during the last four test-list trials gives a significant Words effect; $F(1, 60) = 7.77$, $p < .01$.[1] The Words effect, however, diminished with increased learning-list trials, giving a reliable Words × Trials interaction with $F(1, 60) = 4.71$, $p < .05$. Subsequent t tests for word differences within learning-list conditions were significant for both the S2 and D2 groups ($p < .05$ for $t = 2.26$ and $p < .01$ for $t = 3.53$, respectively) but not for the S6 and D6 groups. The task manipulation produced no reliable differences.

---

[1] The total for the last four trials is taken as the more reliable dependent variable measure since during the first two trials *S*s vary to the extremes in their response strategies. Some push the lever for all the words, while some push it for none of them. This fact, obscured by summary data, is easily seen in the raw scores.

It seems reasonable to conclude that norm-inferred associations of the A-B, C-B type will predict facilitated learning of new responses, although the facilitation grows less with increased practice in the presumed mediator-acquisition stage. The finding that mediation effects occur is, of course, not new in verbal learning. A recent systematic investigation of three-list mediation paradigms (e.g., A-B, C-B, A-C) by Horton and Kjeldergaard (1961) revealed facilitated learning in seven out of eight paradigms studied. Three-list paradigms, however, differ from four-list paradigms using two inferred lists in at least one respect of potential importance for behavioral predictions based upon associative strength, whether strength is derived from normative data or experimental repetition. Storms (1958) has shown that word association frequencies are temporarily distorted when response words are presented to *S*s under some pretext immediately prior to a word association test. This distortion due to recency, manifested in higher response frequencies than would be predicted from norm tables, may account for the fact that attempts to relate normative word association frequency to *amount* of mediated facilitation have generally been unsuccessful (see Jenkins, 1963). Such attempts have usually involved three-list paradigms in which the mediator must be explicitly presented somewhere in the experiment and which may be used by *S*, whether he reports its deliberate use or not, in his learning strategy. Perhaps such a recency effect is partly reduced when a four-list paradigm with two inferred lists is used, since the mediator does not appear explicitly. Then, with word association frequency confounded less with recency effects, the relation between normative frequency and mediated learning should be somewhat more direct. Some evidence for this analysis is found in the present experiment. When the joint probability (relative frequency) of the related word pairs (A-B, C-B) is correlated with the total number of responses to each C word when used as a mediation word less total responses when used as control, the Spearman rho equals .59, which is significant, $t(8) = 2.08$, $p < .05$, by a one-tailed test, even though the words were chosen from a restricted range of associative frequencies.

The finding that mediation effects diminish with additional A-D practice supports the hypothesis that potential mediators are unlearned when they occur in an interference condition. This result is, of course, tentative with two levels of practice and may be restricted to the range of conditions studied. However, it is consonant with the Russell-Storms findings. Although practice on their learning list continued to a criterion, the mediator is postulated not in that stage of the experiment but later, during the test list, where its activation could be reinforced.

The finding that facilitated learning decreases with greater practice on a previous list may be related to the common observation that extensive practice in some learning task or problem solving approach results in difficulty in switching to a new learning strategy. This phenomenon, sometimes called functional fixedness, may result partly from temporary extinction or unlearning of potentially mediating responses in a previous stage of activity.

REFERENCES

BARNES, J. M., & UNDERWOOD, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

HORTON, D. L., & KJELDERGAARD, P. M. An experimental analysis of associative factors in mediated generalization. *Psychol. Monogr.*, 1961, **75**(11, Whole No. 515).

HULL, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.

JENKINS, J. J. Mediated associations: Paradigms and situations. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Ch. 6.

MARTIN, J. G. Mediated transfer in two verbal learning paradigms: Studies in verbal behavior. Report No. 5, 1961, University of Minnesota.

MINK, W. D. Semantic generalization as related to word association. Technical Report No. 17, 1957, University of Minnesota, Contract N8-onr-66216.

POSTMAN, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961. Pp. 152–179.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216.

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

SHIPLEY, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12**, 337–357.

STORMS, L. H. Apparent backward association: A situational effect. *J. exp. Psychol.*, 1958, **55**, 390–395.

(Received August 8, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 17–22

INCUBATION OF ANXIETY AS A FUNCTION OF COGNITIVE DIFFERENTIATION[1]

ELI SALTZ AND DAVID ASDOURIAN

Center for the Study of Cognitive Processes, Wayne State University

The present study tested the hypothesis that the incubation effect represents a spread, rather than an increase in anxiety. 72 albino rats were trained to asymptotic speeds in a runway. 36 were then shocked in the original (gray) goal box, 36 in a different (white) goal box. For Ss shocked in the different box, Ss tested 22 hr. after shock showed significantly (.05 level) greater avoidance behavior than those tested 1 hr. after shock. For Ss shocked in the original goal box, no difference in avoidance occurred between 1-hr. and 22-hr. groups. High differentiating Ss showed significantly (.01 level) less incubation than low differentiators.

Diven (1937) reported that anxiety associated with a specific cue tends to increase in intensity as a function of time, and he called this phenomenon incubation of anxiety. Evidence interpreted as supporting this finding has been reported by Bindra and Cameron (1953), Golin (1961), and Mednick (1957). White (1956) has used this finding in attempting to develop a theoretical account of the development of the neuroses.

The present study has the following purposes:

1. To investigate the hypothesis, suggested by the theoretical position proposed by Saltz (1961), that incubation does not represent an *increase* in anxiety response to trauma associated stimuli as a function of time, but rather that it represents a *spread* of anxiety from trauma associated stimuli to other, previously neutral stimuli.

2. To investigate whether instrumental avoidance responses produced by anxiety display incubation characteristics similar in nature to the physiological responses measured in previous incubation studies.

Saltz (1961) has suggested that S-R systems must be differentiated before a response can be reliably emitted to its appropriate stimulus. A differentiated system has boundaries which prevent intrusion of extraneous responses into the system. With greater overlap between systems (overlap defined operationally in terms of similarity) the intrusion tendency increases. Studies by Saltz (1961) and by Saltz and Riach (1961) indicate that stress breaks down boundaries. This general view leads to the proposition that anxiety, as a response within a bounded system, spreads to overlapping systems with time.

An examination of the previous incubation studies strongly suggests that the reported increases in *intensity* of anxiety with time occurred only in situations where careful controls were absent for distinguishing between an increase in anxiety and a spread of anxiety. Diven's *S*s were presented with verbal stimuli under instructions to free associate to each word. The *S*s were shocked 12 sec. after the presentation of a key stimulus. Since 12 sec. between CS and shock is too long an interval for good anxiety conditioning, it appears reasonable that the real CS was not the original stimulus, but the free associate emitted by *S* immediately prior to shock. Since the free associates were in a chain initiated by the key word, they should in most cases have overlapped with this word in similarity. Consequently, the possibility exists that the increase in PGR to the original stimulus as a function of time, which Diven reports, represents a spread of anxiety from the "real" CS (i.e., the word emitted by *S* immediately before shock) to the originally presented stimulus. The same interpretation is applicable to the Bindra and Cameron (1953) and the Golin (1961) studies, both of which employed a 15-sec. CS-UCS interval, with *S*s emitting responses during the interval.

---

[1] The present study is part of a program of research on differentiation supported by National Science Foundation Grant G 9068.

This interpretation is consistent with Mednick's (1957) finding. Mednick also used a PGR conditioning situation but employed a more adequate 3-sec. CS-UCS interval. She found no incubation to the training CS, but obtained significant incubation to stimuli similar to the training CS.

The present experiment studies the incubation of an instrumental avoidance response in the albino rat. The *S*s were tested for avoidance tendencies either 1 or 22 hr. after shock. The following predictions were made.

1. Anxiety will not increase to the originally conditioned stimulus as a function of time. However, the anxiety response to overlapping stimuli *will* increase as a function of time.

2. Individual differences in *S*'s boundary strength will be related to the spread of anxiety. The weaker an *S*'s boundaries, the greater the spread of anxiety as a function of time.

An index of individual differences in boundary strength was suggested by a study performed by Saltz, Whitman, and Paul (1963) which tested the hypothesis that even such "simple" behavior as running down a straight alley involves differentiation of response systems. In the alley *S* must run, in the goal box he must stop. The stronger the boundary strengths of these competing response systems, the less the stopping response will intrude on alley running, and the faster *S* will run down the alley. Consequently, in the present study the index of individual differences in boundary strength was the asymptotic preshock speed down the alley: greater speed indicated greater boundary strength.

Miller (1961) reports an incubation study in which *S*s ran down an alley to the goal box where they were shocked; half the *S*s were tested for avoidance of the goal box immediately after shock, half 24 hr. later. He points out that there is a greater stimulus change between preshock running and postshock running for the immediate test group than for the 24-hr. test group since for the immediate group the incubation test trial is the second in a sequence (the first trial having been shocked) while for the 24-hr. group the incubation test trial is the first trial in a sequence. In the present study this methodological consideration was taken into account by placing *S*s directly into the shock box (rather than having *S*s find shock at the end of the normal runway locomotion); and also by delaying the short time interval group 1 hr. after shock, rather than testing them immediately after shock.

METHOD

Subjects.—The *S*s were male albino rats obtained from the Holtzman Company in Madison, Wisconsin. The 12 *S*s in the control group and the 72 *S*s in the four experimental groups were approximately 85 days old on the first day of training.

Apparatus.—A straight alley was used that was 90 in. long, 5 in. high, and 3½ in. wide. Guillotine doors were placed at both ends of the alley and served to divide the area into a start box 9 in. long, a runway 72 in. long, and a goal box 9 in. long. The apparatus was made of wood with the roof constructed with ¼-in. hardware cloth. The floor of the start box and the runway were made of wood, but the floor of the goal box consisted of a grid of ⅛-in. diameter brass rods placed ½ in. apart from one another and was wired for foot shock. The foot shock was delivered through a 100,000-ohm fixed resistor in series with the animal. A variac was used to step down the 60-cycle 115-v. house current to 90 v. Every *S* received 90 v. during the shock trials. The whole apparatus was painted a medium gray. The goal box was removable. A second goal box was also used which was essentially identical to the gray goal box except that it was painted white. The gray and white goal boxes were interchanged as described below.

The opening of the start-box door started two Standard Electric timers calibrated in seconds and .01 sec. Just beyond the start-box door were two metal plates wired to a capacitance operated relay. As soon as *S*s left the start box they made contact with both plates and stopped the first clock. Two more plates were positioned in front of the goal-box door and the second clock was stopped when *S*s stepped on these two plates. The reward used during acquisition consisted of two food pellets, weighing 45 mg.

Pretraining.—Days 1 and 2: The start-box doors were kept open and *S*s were put into the alleyway in groups of four or five and were permitted to explore the apparatus for 10 min. After exploring the apparatus on Day 2 the animals were returned to their home cages and deprived of food.

On Day 3 following approximately 22 hr. of deprivation, each *S* was placed directly into the gray goal box and was permitted to eat food pellets for 3 min. After being fed, all of the animals were returned to their home cages and permitted to eat ad libitum for 1 hr. At the end of 1 hr. of feeding, all food was removed.

Acquisition.—During acquisition, *S*s learned to run to the gray goal box and eat food pellets. All *S*s were given five distributed trials per day; each *S* had a trial before any *S* ran his next trial. The *S*s were run in random order on any given trial. Each was put into the start box and goal-box doors were opened. As soon as *S* passed through a door the door was closed. The *S*s were given a maximum of 5 min. to traverse the runway, and a maximum of 2 min. to eat the pellets in the goal box. The *S*s were run on acquisition until they reached asymptotic speed. The measure of speed that was used was the interval of time between raising both doors and stopping the second clock. Asymptotes were reached during the eighth or ninth day (40 to 45 trials).

Testing.—The *S*s were assigned to one of six groups. These groups were: (*a*) no shock in gray goal box—test after 1 hr.; (*b*) no shock in gray goal box—test after 22 hr.; (*c*) shock in gray goal box—test after 1 hr.; (*d*) shock in gray goal box—test after 22 hr.; (*e*) shock in white goal box—test after 1 hr.; (*f*) shock in white goal box—test after 22 hr. There were 6 *S*s in each of the two nonshock control groups. The four shock conditions were run with 18 *S*s in each group. The *S*s were matched on the basis of their last day's running times and then assigned to the various groups. Six replications were run in completing the study. All of the *S*s under all conditions were deprived of food approximately 22 hr. prior to the beginning of the test procedure.

Control groups.—Each *S* in the control groups was put into the gray goal box and detained there for 1 min. After all of the control *S*s had been detained, the procedure was immediately repeated for these *S*s. The 1-hr. controls were returned to their home cages for 1 hr. and then were given five acquisition trials to a gray goal box. The 22-hr. controls were returned to their home cages and fed after 1 hr. The next day these *S*s were given five acquisition trials in the gray goal box.

Experimental groups.—The *S*s shocked in the gray goal box and run to the gray goal box were treated in the following way:

1. All *S*s were given five 3-sec. shocks to their paws with 10-sec. intervals between shocks. After all of the *S*s had been shocked, each was returned to the goal box and the shock procedure was repeated. There was no food present in the goal box during the shock trials.

2. The 1-hr. and the 22-hr. delay groups were returned to their home cages but not fed. After 1 hr., the 1-hr. group was run on five acquisition trials.

3. The 22-hr. group was fed 1 hr. after being returned to the home cages. This group was run the next day on acquisition approximately 22 hr. after having been fed.

The *S*s shocked in the white goal box and

TABLE 1

Mean Running Time on Last Training Day and Running Time on First Test Trial for Control Groups and Combined Experimental Groups

Groups	Last Training Day		First Test Trial	
	M	*SD*	*M*	*SD*
Control				
1 hr.	1.62	.44	1.66	.36
22 hr.	1.60	.35	1.57	.47
Experimental	1.93	.45	72.13	77.56

run to the gray goal box were treated exactly the same way as *S*s shocked in the gray goal box. The control *S*s and the experimental *S*s were all given a maximum of 5 min. to traverse the runway and 2 min. to eat the food.

Results

Both running time on the first trial after shock and mean running time over the five trials following shock were analyzed. Since these results parallel each other closely, only the data for the first trial after shock are presented.

The results of the nonshocked control group, summarized in Table 1, indicate increased running times obtained in the experimental conditions are due to shock produced avoidance rather than extinction of the running response as a function of being placed in the goal box and not rewarded (i.e., latent extinction). The mean speed on the last day of training is almost identical to the speed on the first test trial for both 1- and 22-hr. groups. Analysis of variance (Lindquist, 1953) indicates no significant differences for main effects or interaction, all three *F* ratios being close to 1.00. The results for the 72 experimental *S*s are in sharp contrast to that of the controls. As shown in Table 1, *S*s ran approximately 35 times as fast during the last training trial as on the first test trial. Not a single *S* ran faster on the first test trial than his own mean performance on the last training day.

The mean running times for the four basic shock groups are presented in Table 2. The first prediction stated that anxiety will *not* increase to the originally conditioned stimulus as a function of time. Consistent with this prediction, there is almost no difference in avoidance between the 1- and the 22-hr. groups which were shocked in a gray box and then tested in a gray alley. The first prediction also stated that the anxiety response *will* increase as a function of time to stimuli which *overlap* with the originally conditioned stimulus. The data are also consistent with this aspect of the prediction: For *S*s shocked in a white box and tested in a gray alley, there is clear-cut evidence for an incubation effect—avoidance is substantially greater 22 hr. after shock than 1 hr. after shock. The variance for the various groups proved to be heterogeneous, and the distributions had extremely different shapes making parametric statistics inapplicable.

The nonparametric statistics reported below were chosen on the basis that incubation was predicted to be a function of degree of boundary strength involved in differentiation,

TABLE 2

Mean Running Time on First Test Trial

Shock Cond.	Interval between Shock and Test			
	1 Hr.		22 Hr.	
	M	*SD*	*M*	*SD*
White	48.03	72.98	94.19	105.93
Gray	71.21	44.94	75.08	75.97

and that *S*s in the 1- and 22-hr. groups were matched on degree of boundary strength (using mean running time on the last day of training as an index of boundary strength). A Wilcoxon matched ranks test was used to evaluate the difference in running time for 1-hr. vs. 22-hr. delay after shock. With 18 matched pairs, a Wilcoxon *T* of 40 or less is required for significance at the 5% level. For *S*s shocked in the white box and tested in the gray, the *T* for the difference between 1- and 22-hr. delay was 35. For *S*s shocked in the gray box tested in the gray, $T = 78$; in 10 of the 18 pairs the 1-hr. *S*s actually showed greater avoidance than the 22-hr. *S*s. The results suggest, therefore, that incubation of anxiety consists of a spread of anxiety from trauma associates stimulus systems to other systems as a function of time.

To test the second prediction, that incubation is related to degree of boundary strength, rho's were completed between the index of boundary strength for each pair (i.e., mean running time on the last training day) and the difference in running time for each 1-hr. and 22-hr. pair of *S*s. For *S*s shocked in a white box and tested in a gray, rho = +.68. Based on 18 pairs of *S*s, this correlation is significant at well beyond the 1% level. For *S*s shocked in a gray box and tested in gray, rho was −.18, which is far from significant. These results also support the writers' predictions.

Finally, it will be noted in Table 2 that the mean avoidance for the 22-hr. group shocked in the white box exceeds that for the 22-hr. group shocked in gray. While no specific prediction was made on this point, it should be apparent that the spread of anxiety notion would demand that the anxiety response to the test stimuli for the white-shock *S*s should not significantly exceed that of the gray-shock *S*s. If the obtained difference were significant, it would suggest that more anxiety spread from the trauma associated system to the previously neutral system then was originally present in the trauma associated system. A test of this difference was performed by matching 22-hr. white and 22-hr. gray *S*s on mean running time on the last training day, and computing a Wilcoxon *T*. This proved to be far from significant, only 9 of the 18 pairs producing a difference indicating greater avoidance for the white-shock *S*s.

DISCUSSION

The following conclusions are suggested:

1. Incubation appears to be a phenomenon involving *spread* of anxiety rather than an absolute increase in *S*'s anxiety level. The results of the present study, involving an instrumental avoidance response, are consistent with Mednick's (1957) results involving PGR. In both studies, when the critical cue was identical between stress and test, no incubation occurred. Only when the trauma associated cue and the test cue were similar but not identical did the anxiety reaction to the test cue increase with time. These results support the writers' analysis of the Diven (1937), Bindra and Cameron (1953), and Golin (1961) studies discussed earlier in this paper.

2. Incubation is a function of the boundary strength differentiating the relevant S-R systems. The .68 correlation obtained between the index of boundary strength and incubation for the white-shock group is surprisingly high when one considers that this index is contaminated by other variables such as the physical strength of *S*s.

3. Contrary to Miller's (1961) position on this issue, the evidence suggests that incubation of an instrumental avoidance response can occur.

The writers feel that the present data present some difficulty in interpretation for more traditional S-R formulations. If by "generalization" one means merely the empirical fact that a response trained to one stimulus may be evoked by similar stimuli, then the present study suggests that generalization of anxiety increases as a function of time. If "generalization" is a hypothetical construct which *predicts* transfer from a training stimulus to other stimuli, traditional S-R theory appears inadequate for indicating the characteristics of the hypothetical construct necessary for predicting the obtained data. Nor is it clear how traditional S-R theory would conceptualize the individual differences in differentiation which appear so critical to incubation in the present study.

REFERENCES

BINDRA, D., & CAMERON, L. Changes in experimentally produced anxiety with the passage of time. *J. exp. Psychol.*, 1953, **45**, 197–203.

DIVEN, K. E. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.

GOLIN, S. Incubation effect: Role of awareness in an immediate versus delayed test of conditioned emotionality. *J. abnorm. soc. Psychol.*, 1961, **63**, 534–539.

LINDQUIST, E. F. *Design and analysis of experiments.* Boston: Houghton Mifflin, 1953.

MEDNICK, M. T. Mediated generalization and the incubation effect as a function of manifest anxiety. *J. abnorm. soc. Psychol.*, 1957, **55**, 315–321.

MILLER, N. E. Some recent studies of conflict behavior and drugs. *Amer. Psychologist*, 1961, **16**, 12–24.

SALTZ, E. The effect of induced stress on free associations. *J. abnorm. soc. Psychol.*, 1961, **62**, 161–164.

SALTZ, E., & RIACH, W. J. The effect of stress on stimulus differentiation. *J. exp. Psychol.*, 1961, **62**, 588–593.

SALTZ, E., WHITMAN, R. N., & PAUL, C. Performance in the runway as a function of stimulus-differentiation. *Amer. J. Psychol.*, 1963, in press.

WHITE, R. W. *The abnormal personality.* New York: Ronald Press, 1956.

(Received August 8, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 23–29

EFFECT OF "SUBLIMINAL" TONES UPON THE JUDGMENT OF LOUDNESS[1]

WILLIAM BEVAN AND JOAN FAYE PRITCHARD

Kansas State University

A pilot study and 3 experiments, involving a total of 144 Ss, were conducted to determine if a "subliminal" anchor effect, already reported for the sensory dimension associated with mild electric shock, could be demonstrated also for the dimension of loudness. After some manipulation of the experimental conditions, it was found that the introduction of a tone below the clearly audible level was accompanied by an elevation in the judged magnitude of series stimuli clearly above this level. It is noteworthy that the anchor effect was finally produced by increasing the scalar distance between the series and the anchor-designate, although this involved decreasing the physical intensity of the anchor and making it less likely to be heard. Implications of the results for adaptation level theory are suggested.

In a recent paper, Black and Bevan (1960) report that "subliminal" stimulation, introduced without *S*'s knowledge, may effect an increase in the judged intensity of supraliminal series stimuli. Their procedure involved presenting an experimental group electric shocks below the range of threshold sensitivity and interpolated at the midpoint of the temporal interval between successive mild shocks used as series stimuli. Comparison of the responses of this group with those of a control group not presented the subliminal shocks revealed two differences worthy of note: First, the series shocks were judged consistently more intense by the experimental *S*s. Second, the slope of the psychophysical function for the experimental group was flatter than that for the control, the shift in the experimental judgments being greater for the weaker members of the series than for the stronger. These results suggest that extremely subtle stimulation may influence the magnitude of psychophysical judgments and that this influence resembles that produced by clearly perceptible anchors below the series.

The Bevan and Black data have been confirmed by Goldstone and his colleagues (Goldstone, Goldfarb, Strong, & Russell, 1962), although the latter found it necessary to modify procedure so that each interpolated subliminal stimulus was temporally closer to the series member it preceded.

The purpose of the present series of experiments was to further establish the Black-Bevan effect as a bona fide anchor effect by demonstrating it for a second sensory dimension. Since electric shock is a noxious stimulus with broad biological significance for *S*, it is conceivable that the subliminal anchor effect may be peculiar to this and similar dimensions. Therefore, the dimension chosen for the experiments herein described was loudness.

---

[1] The authors are indebted to Josephine Baker and Lonnie D. Whitehead, who performed the pilot experiment as part of an advanced course in experimental psychology. They are also grateful to Gherry Harding and Jerome Walker, who jointly served as *E*s for Exp. I while participants in a National Science Foundation Undergraduate Research Participation Program during the summer of 1960, and James Haines, who helped with data collection for Exp. II. The junior author, Joan Faye Pritchard, served as *E* for Exp. II and III. Experiments II and III were performed under Contract Nonr 3624(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research.

The *E* who undertakes to work with subliminal stimuli faces a number of difficult methodological problems. For his experiment to be subliminal in the strict sense of the word stimuli must lie below threshold on each trial on which they are presented. But threshold is an arbitrarily defined statistical concept. Stimuli which lie outside definitional criteria may still be reported and estimates of threshold will vary with the method and conditions under which they are obtained. He may, as we have in the experiments reported below, reference his identification of subliminal intensities to a particular method (here the ascending method of limits) and exclude all *S*s who on as much as a single trial say or otherwise verbally indicate that they are aware of the presence of the subliminal stimulus. But the threshold will also systematically shift under the influence of a variety of variables. For example, Bricker and Chapanis (1953) have reported that while correct guesses of tachistoscopically-presented words may occur only a chance number of times on an initial trial, they will occur better than chance on the second trial, and Boardman (1957), in a study of preword recognition hypotheses, reported that response criteria change with practice. Furthermore, *E* must recognize, as Eriksen (1958) insists, that the perceptible is not co-extensive with the verbalizable. This confronts him with the murky problem of levels of awareness. But the question as to whether or not *S* is in any sense *really* aware of the subliminal stimulus is largely a metapsychological issue and is not crucial for studies such as this one. Rather, what is of concern to us here is whether or not anchor stimuli, subtle beyond the usual criteria of recognition, can still exert an influence upon ordinary psychophysical judgments.

EXPERIMENT I

At the time that the Black-Bevan experiment was nearing completion, a pilot experiment involving the judgment of loudness was carried out. This consisted of a counterbalanced design in which all *S*s judged the series members under both control (no-anchor) and experimental (anchor) conditions. For half, the control condition occurred on the initial session, with the experimental condition being presented several days later; for the other half, the order was reversed. Tones to be judged were presented through head phones with ambient white noise of moderate intensity present to mask incidental sounds that might serve as anchors above the series.

The results plotted as two functions separated by a fraction of a category-unit's difference at the low end of the series and converging on a common value at the upper end of the series. However, the relative position of the two curves was opposite to expectation; the judgments in the presence of the subliminal stimulus were *less* intense than those of the control condition. While the between-conditions source of variance was not significant, the difference in slopes (Conditions × Stimuli interaction) was highly significant ($p < .001$).

Since a preliminary experiment with shock had shown already that order of presentation influences anchor effectiveness, it was conjectured that these results might reflect the operation of this variable. Accordingly, Exp. I is a repetition of the loudness experiment using the simple two independent-groups design.

Method

Subjects.—The *S*s were 30 undergraduate students enrolled in introductory psychology: 7 men and 8 women made up the control group while 8 men and 7 women constituted the experimental group. None had previous experience in psychophysical experiments.

Apparatus.—The stimulus tones were generated by a Hewlett-Packard audio oscillator, Model 200CD and presented binaurally to *S* through Telephonics Type TDH-39 high fidelity earphones. The intensity of the stimuli (1,000-cps tones) was set manually on each trial by the *E* by means of a Hewlett-Packard Model 350B attenuator. The duration of each presentation as well as the intervals between presentations was controlled by two Hunter Model 111 C Golden Silence interval timers appropriately connected for recycling. The elimination of clicks associated with onset and offset of the tones, as well as control of their rise and decay times, was accomplished through use of a Grason-Stadler electronic switch. The output of a General Radio Type 1890A Random Noise generator, amplified through a Bell Model 2122-C amplifier and presented through an Oxford 8-in. speaker provided the ambient sound screen.

Procedure.—Each *S* was tested individually. The first step in the experimental procedure involved determining his loudness threshold in the presence of a continuous noise level of 71 db. A variation of the method of limits was used, the threshold being taken as the median of 15 momentary estimates obtained with an ascending series alone.

The *S* was next instructed in the judgment of sensory stimuli using the rating scale version of the absolute method. He was given nine categories, varying from *very, very soft* through *medium* to *very, very loud,* but was told to use as many categories as possible and to add categories when appropriate. Finally, it was explained that *loud* and *soft* had reference only to the stimuli he would be presented during the test session.

Each *S* received 20 presentations of each of five series intensities, with the order of presentation random for five successive blocks of 20 trials each. The physical intensities used were set with reference to the individual *S*'s absolute threshold. The weakest member of the series was 5 db. above this value, with the additional stimuli represented by successive increments of 5 db. The subliminal anchor was set at 5 db. below threshold and was interpolated without *S*'s knowledge. Care was taken to eliminate any member from the experimental group who assigned a category value on one or more presentations of the *subliminal* stimulus or who either voluntarily or upon questioning at the end of the test session gave evidence of being aware of its presence. The series stimuli were presented at intervals of 10 sec. for a duration of 1 sec. The subliminal stimulus was introduced at the midpoint of the presentation interval, i.e., 5 sec. after the onset of each series stimulus. Medians of the category judgments made for the several series stimuli by individual *S*s served as data for statistical analysis. Psychophysical functions for the two groups were constructed from the means of these medians.

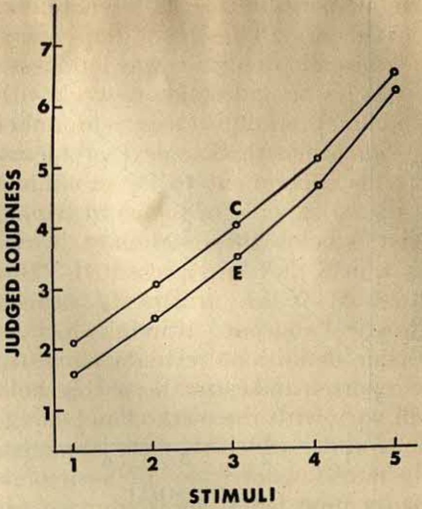


FIG. 1. Judged loudness of tones for experimental and control groups of Exp. I.

Results

The results of Exp. I are seen in Fig. 1. It may be noted that there are two nonoverlapping functions with maximum separation at the low end of the series. Statistical evaluation by means of a trend test indicates the separation of the two curves to be reliable, $F_{\text{between groups}}(1, 84) = 52.22$, $p < .001$, while it fails to confirm the asymmetrical anchor effect, $F_{\text{between slopes}}(1, 84) = 1.32$, $p > .05$. These data are in line with those of

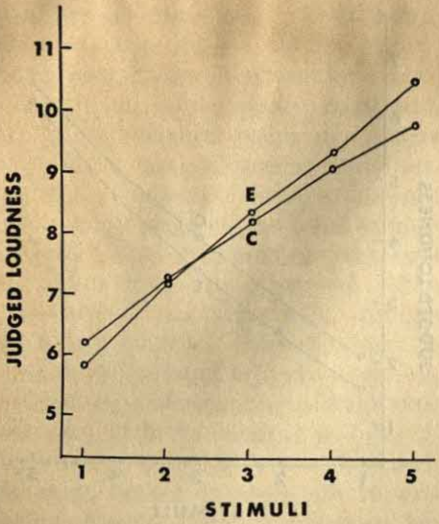


FIG. 2. Judged loudness of tones for the control and experimental groups of Exp. II.

the pilot experiment. The judgments of the experimental group, contrary to expectation, are *less* intense than those of the control, but, since the design involved independent groups, cannot be attributed to an interaction involving order effects. Two other possibilities suggest themselves. It is conceivable that the masking noise, instead of screening out incidental anchors, itself constitutes an anchor above the series, interacting with the experimental conditions to produce the obtained results. On the other hand, the subliminal anchor, being relatively impotent and close to the series, could have effected *assimilation* rather than the more frequently-observed *contrast* anchor effect. Therefore, Exp. II was undertaken to check these possibilities.

EXPERIMENT II

Experiment II was essentially a repetition of Exp. I under circumstances which allowed the omission of the ambient white noise. This was accomplished by conducting the test session in a specially constructed test chamber, capable of attenuating extraneous noises by approximately 60 db., with the *S* wearing headphones mounted in ear muffs capable of the further dampening of outside sound by 45–60 db.

Method

Subjects.—Eighty introductory psychology students, 20 men and 20 women in the control and experimental groups, respectively, participated as *S*s.

Apparatus.—The apparatus was that of Exp. I except that the Telephonics Type TDH-39 phones were replaced by Willson Sound Barrier earmuffs containing Stromberg Carlson high-fidelity phones.

Procedure.—The procedure of Exp. II was identical with that of Exp. I except that the physical intensities of all stimuli were the same for all *S*s. The series intensities were 26, 31, 36, 41, and 46 db. SPL, respectively. The subliminal stimulus was 16 db. SPL. Each *S* in the experimental group was carefully observed while being tested and was interrogated at the end of the session. Twenty-five evidenced knowledge of the presence of the subliminal input and were disqualified as *S*s. These are in excess of the 80 identified above as *S*s.

Results

The elimination of the sound screen and the insulation of *S* from extraneous noise yielded data clearly different from those of the pilot study and Exp. I (cf. Fig. 2). It thus seems reasonable to attribute the reversal in the relative positions of the experimental and control curves obtained in these earlier investigations to the anchor-like intrusion of the masking noise. Meanwhile, the results are not definitive confirmation of the proposition that subliminal stimulation may behave like an anchor below the series. Although the E group values are greater than corresponding C group values for the three loudest members of the series, the curves seen in Fig. 2 intersect at a value approximately equal to that of Stimulus 2 and the reversal for the weakest series member suggests a slight assimilation effect. Analysis of variance of the data of Exp. II indicates no reliable between-groups difference, $F_{\text{between groups}}(1, 76) = 0.33$, $p > .05$, but a significant difference in slopes, $F_{\text{Groups} \times \text{Stimuli}}(4, 304) = 4.75$, $p < .01$.

EXPERIMENT III

The purpose of Exp. III was twofold: to seek a more effective set of stimulus conditions for the production of the subliminal anchor effect, and, as part of this, to explore the result of changing the temporal position of the anchor within the presentation interval. Goldstone, Goldfarb, Strong, and Russell (1962) indicate that when the subliminal anchor effect for shock could not be obtained with interpolation at the midpoint of the presentation interval, it was possible to evoke it by moving the anchor back so that it preceded each series member by 2 rather than 5 sec. Similarly, Maruyama (1957), in a study of intermodal relations, interpolated a tone either immediately following the standard visual stimulus or just prior to the comparison stimulus and found that the tone influenced the brightness judgment only when it occurred in the latter temporal position. Accordingly, it was decided to examine the subliminal anchor effect with the anchor at several positions within the presentation interval. Since inspection of Fig. 2 suggests a weak and inconsistent anchor effect, it was assumed that this effect might be enhanced by increasing the scalar distance between series and anchor magnitudes.

Method

Subjects.—Three groups of 20 *S*s each, 10 men and 10 women, represented the experimental condition. The control group of Exp. II was used to provide control data for this experiment. As in the case of Exp. II, *S*s indicating cognizance of the subliminal

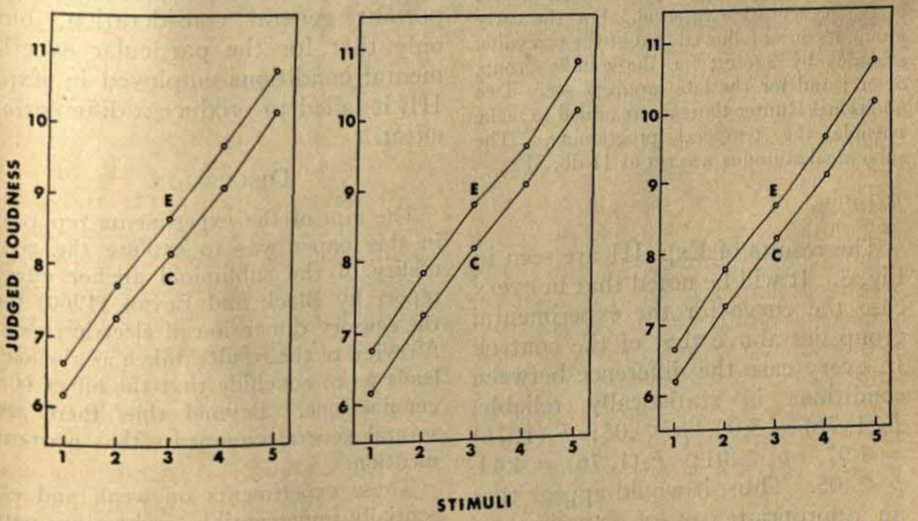


FIG. 3. Judged loudness for the groups of Exp. III. (The left-hand graph presents the data obtained when the anchor occurred early—3 sec. after onset of the preceding stimulus—in the presentation interval; the center graph, the data when the anchor occurred at the midpoint of the interval—5 sec. after onset of the preceding stimulus; and the right-hand graph, the data when the anchor occurred late in the presentation interval—7 sec. after onset.)

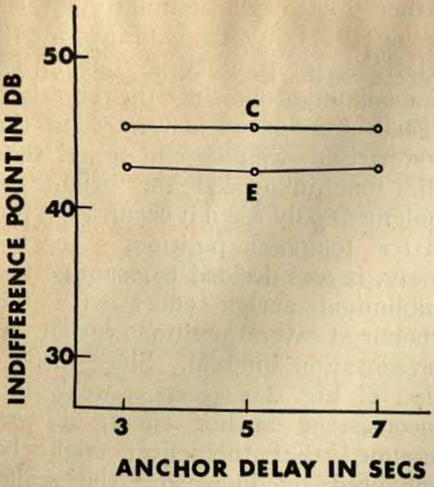


FIG. 4. Magnitude of the anchor effect as a function of the position of the anchor in the presentation interval. (The medium heaviness judgment is represented by the ordinate; the temporal proximity of the anchor to the preceding series stimulus on the abscissa.)

inputs were eliminated and are in excess of the 60 designated as *S*s.

Procedure.—The subliminal stimulus was introduced early in, at the middle of, or late in the presentation interval. For the early group, its onset followed that of the preceding stimulus by 3 sec.; for the middle group, 5 sec.; and for the late group, 7 sec. Two additional Hunter timers were added to make possible the temporal programing. The subliminal stimulus was set at 11 db. SPL.

Results

The results of Exp. III are seen in Fig. 3. It will be noted that in every case the curve for the experimental group lies above that of the control. In every case the difference between conditions is statistically reliable, $F_3(1, 76) = 5.09$, $p < .05$; $F_5(1, 76) = 7.97$, $p < .01$; $F_7(1, 76) = 4.44$, $p < .05$. Thus, it would appear that an appropriate set of experimental conditions had obtained and the subliminal anchor effect observed by Black and Bevan for shock was confirmed for loudness. Similar support for the generality of the subliminal anchor effect is provided by a recent study by Boardman and Goldstone (1962) on the judgment of size. Using a visual-masking technique with tachistoscopically-presented discs, these *E*s supplemented the psychophysical series with an anchor disc either larger than or smaller than the series members. Although the differences were small, the judgments for the group receiving the small anchor were consistently greater for the experimental than for the control condition; while those for the group receiving the large disc were, as expected, less than the control judgments.

Figure 4 is a graph of the adaptation levels (the stimulus magnitudes corresponding to the judgment of *medium*) for the experimental groups of Exp. III expressed in db. It is clear from the zero slope that the subliminal anchor effect did not differ over the range of interpolation times used. This, of course, does not mean that interpolation time is not an important general consideration, but only that for the particular experimental conditions employed in Exp. III, it failed to produce a differential effect.

DISCUSSION

The aim of the experiments reported in this paper was to explore the generality of the subliminal anchor effect report by Black and Bevan (1960) for the sensory dimension of electric shock. A review of the results, taken as a whole, leads us to conclude that the effect is a genuine one. Beyond this, there are several general comments that warrant mention.

These experiments on weak and essentially imperceptible anchors are part of a larger program of studies dealing with the problem of relevance in sensory judgment. More specifically, this program is directed toward identifying the properties that identify stimuli that are pooled by the judge and therefore determine the scalar value of an adaptation level in contrast to those which exert no influence upon the judgmental norm. The present results are consistent with the Black and Bevan conclusion that the classically-defined absolute threshold need not be a limiting condition in pooling.

Two characteristics of the Black and Bevan results support the conclusion that the subliminal anchor effect is a genuine anchor effect: (*a*) the judgments of series stimuli with the anchor interpolated were significantly greater than when it was absent; and (*b*) the subliminal anchor effect is greatest for stimuli which are at the low end of the series and thus also nearest the position of the anchor in the stimulus scale. Comparison of Exp. II and III adds a third characteristic. (*c*) The subliminal effect was enhanced by changing the intensity of the subliminal stimulus. This involved *increasing* the scalar distance between the anchor designate and the series members while at the same time *reducing* its physical magnitude and taking it *farther* below initial threshold.

REFERENCES

Black, R. W., & Bevan, W. The effect of subliminal shock upon the judged intensity of weak shock. *Amer. J. Psychol.*, 1960, **73**, 262–267.

Boardman, W. K. Utilization of word structures in prerecognition responses. *J. Pers.*, 1957, **25**, 672–685.

Boardman, W. K., & Goldstone, S. Effects of subliminal anchors upon judgments of size. *Percept. mot. Skills*, 1962, **14**, 475–482.

Bricker, P. D., & Chapanis, A. Do incorrectly perceived tachistoscopically-presented stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

Eriksen, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958. Pp. 169–226.

Goldstone, G., Goldfarb, J., Strong, J., & Russell, J. Replication: The effect of subliminal shock upon the judged intensity of weak shock. *Percept. mot. Skills*, 1962, **14**, 222.

Maruyama, K. The effect of tone on the successive comparison of brightness. *Tohoku psychol. Folia*, 1957, **15**, 55–70.

(Received August 13, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 30–36

RESPONSE EQUIVOCATION AND REACTION TIME[1]

ROBERT E. MORIN AND BERT FORRIN

Kent State University *University of Washington*

This study was designed to assess the relative contributions of transmitted and response information to choice reaction time (RT) in a task involving a 1-many mapping of stimuli into responses. 35 Ss served in 5 independent conditions which varied in terms of number of stimuli displayed (2 or 4) and number of responses per stimulus (1, 2, or 4). In conditions with more than 1 response per stimulus, Ss were instructed to select randomly from available correct response alternatives; this procedure effected a partial unconfounding of response and transmitted information. Ss were given 576 RT trials distributed over 3 days of practice. Differences in RT between conditions were more clearly associated with differences in response uncertainty than with differences in amount of information transmitted.

In 1952 Hick advanced and presented evidence in support of the proposition that choice reaction time (RT) is a positive linear function of information transmitted. Data consistent with this view have been obtained by a number of investigators (Brainard, Irby, Fitts, & Alluisi, 1962; Bricker, 1955; Hyman, 1953; Merkel, 1885).

At least two major issues arise with respect to the evaluation of Hick's position. The first concerns the recent evidence demonstrating the independence of RT and the informational properties of the task under conditions of extended practice (Mowbray & Rhoades, 1959; Seibel, 1959) and high S-R compatibility (Leonard, 1959). These studies have challenged the universality of the suggested relationship between RT and transmitted information and have indicated the need for research designed to define boundary conditions.

A second issue, and one with which this study is more intimately concerned, involves the investigation of the nature of the RT transmitted information function unconfounded by the effects of stimulus and response uncertainty. For the case of a one-to-one mapping of stimuli into responses, errorless, or near-errorless, performance results in the inextricable confounding of measures of stimulus, response, and transmitted information. This problem exists in most studies demonstrating a positive relation between transmitted information and choice RT. Thus, though they are consistent with Hick's position, they permit alternative views emphasizing the role of stimulus or response uncertainty in information processing behavior.

The problem of confounding arises in part because transmitted information, by definition, can be no larger than either stimulus or response information. A partial solution is to be found by permitting stimulus information to exceed transmitted information. Two experiments (Archer, 1954; Morin, Forrin, & Archer, 1961) accomplished this end by mapping more than a single stimulus into each response (a many-one mapping) thereby making irrelevant some of

---

[1] This study was supported by National Science Foundation Grant G-14292 awarded to the University of Texas and by National Science Foundation Grant G-19173 awarded to the University of Arizona.

the stimulus information. In these studies the presence of irrelevant stimulus information did not serve to degrade performance, a finding consistent with the view that stimulus uncertainty per se exerts no major influence on RT.

Hick's own research provides the only data collected for the explicit purpose of assessing separately the effects of response uncertainty and of transmitted information on RT. In order to at least partially unconfound these variables, he required *S*s to respond at rates faster than those at which they could maintain errorless performance. Since transmitted information is essentially a nonmetric correlation between stimulus and response, the effect of nonsystematic errors was to reduce amount of information transmitted. The distribution of responses, however, and hence response information, was relatively unaffected. Under these conditions choice RT remained proportional to transmitted information.

The present study represents an attempt to isolate the relative contributions of transmitted information and response uncertainty to choice RT without forcing *S*s to make errors. Response and transmitted information were unconfounded by introducing response equivocation. This was accomplished by mapping each stimulus into more than a single correct response (a one-many mapping). Thus, responses were not completely predictable from stimuli even with errorless performance. The *S*s were instructed to attempt to select randomly from among the available correct response alternatives to each stimulus displayed. Since, in the present case, the upper bound on transmitted information was set by stimulus uncertainty, the procedure permitted response information to exceed transmitted information in some experimental conditions.

METHOD

Subjects.—The *S*s were 35 students from classes in elementary psychology at the University of Texas. Participation in the experiment satisfied a course requirement. Seven *S*s were assigned at random to each of five conditions.

Apparatus.—Each *S*, tested individually, was seated approximately 2 ft. in front of a milk-glass viewing screen behind which an automatic 2 × 2 in. slide projector was located. The dimensions of the projected image, a geometric form, were 2 × 2 in. Reaction time was measured to the nearest .01 sec. by a Standard Electric timer which started when a stimulus slide was projected. A verbal response by *S* activated a voice key which simultaneously stopped the timer, terminated the projected image, and initiated an automatic slide changing sequence. Elapsed time between the termination of one slide and the onset of the succeeding slide was 2.1 sec. During this interval *E* recorded *S*'s RT and response.

Experimental conditions.—All *S*s were required to associate men's names with simple geometric forms. The stimuli employed were a circle, a square, a triangle, and a cross. The men's names were BEAU, BEN, BERT, BILL, BING, BOB, BUCK, and BUD. The *S*s were instructed to regard the forms as "team names" and the men as "team members." Five conditions varied in terms of the number of different stimuli (teams) used and the number of responses per stimulus (members per team).

In all conditions the stimulus series was constructed to make the included stimuli equally likely and to equate the first-order transitional probabilities. The amount of information per stimulus in a condition was, therefore, the logarithm to the base two of the number of possible stimuli. The maximum possible information per response was similarly the logarithm to the base two of the number of available response alternatives. Since amount of information transmitted cannot exceed either stimulus or response uncertainty, the smaller of these two measures in each condition established an upper bound on information transmitted. The experimental conditions and their informational properties are summarized in Table 1.

TABLE 1

EXPERIMENTAL CONDITIONS

Cond.	Stimuli	Responses per Stimulus	Information		
			Stimulus	Maximum Transmitted	Maximum Response
I	2	1	1	1	1
II	2	2	1	1	2
III	2	4	1	1	3
IV	4	1	2	2	2
V	4	2	2	2	3

In Cond. I two stimuli were used, the particular two being randomly determined for each *S*. A single response, again randomly determined, was associated with each stimulus. For example, one *S* in Cond. I responded BOB to a square and BUD to a cross. There was thus 1 bit of information per stimulus and, with errorless performance, 1 bit per response in Cond. I. Maximum transmitted information was similarly 1 bit.

Condition II differed from Cond. I in that there were two correct responses to each of two stimuli, four responses in all. The stimuli used and the particular stimulus-response combinations employed were randomly determined for each *S*. To illustrate, for one *S* the stimuli were a square and a triangle; the correct response to the square was either BUCK or BERT, the correct response to the triangle, either BEN or BOB. The *S*s were instructed to be unsystematic in choosing from the correct response alternatives. If an *S* in Cond. II were successful in randomly distributing his response choices, his output would be 2 bits of information per response. Even were response preferences and sequential dependencies to reduce response information below its maximum possible value, it would still be anticipated that amount of information per response would exceed stimulus (1 bit) and maximum transmitted information (1 bit).

In Cond. III there were four correct response alternatives for each of the two stimuli, or eight responses in all. Stimulus and maximum transmitted information remained at 1 bit while maximum information per response was increased to 3 bits. In Cond. IV and V stimulus uncertainty was increased to 2 bits (four stimuli). With one response per stimulus in Cond. IV and two responses per stimulus in Cond. V maximum possible response information was 2 and 3 bits, respectively. Maximum transmitted information, attainable with errorless performance, was 2 bits in both conditions.

Pretraining.—The associations between the geometric forms and the response names were learned in a 25-min. pretraining period. Tape recorded instructions guided *S*'s study of the materials to be learned, afforded opportunity for self-testing, and prepared *S* for the experimental sessions to follow. The practice and self-tests were designed to minimize the acquisition of response preferences.

Procedure.—Each *S* appeared for three experimental sessions at 48-hr. intervals. The first session, which included the pretraining period, lasted approximately 1 hr.; the two subsequent sessions each lasted approximately ½ hr. Each session consisted of 6 blocks of 32 trials. Blocks were separated by 2-min. rest periods. The stimulus series within a block was predetermined as previously indicated. Four unique series, each 32 slides in length, were generated for the two-stimulus and four-stimulus conditions. The sequence in which the series were presented was randomly determined for each *S* with the restrictions that each series appear at least four times in the 18 blocks and that no series follow itself. To minimize the identification of stimulus series, three randomly selected forms were inserted at the beginning of each block.

During test trials a new stimulus was presented 2.1 sec. after the *S* had responded to the previous stimulus. A rapid self-paced schedule was employed to minimize the likelihood that *S* would decide which response to make to a stimulus even before that stimulus appeared. The instructions to *S* emphasized both speed and accuracy. A noncorrection procedure was used since the voice key did not respond differentially to correct and incorrect responses.

RESULTS

In Fig. 1 mean RTs for the experimental conditions are plotted as a function of blocks of 32 trials. Times for both correct and incorrect responses were used in the computation means. In all conditions there was some improvement in performance with practice, the greatest improvement occurring in the initially more difficult conditions. The relative

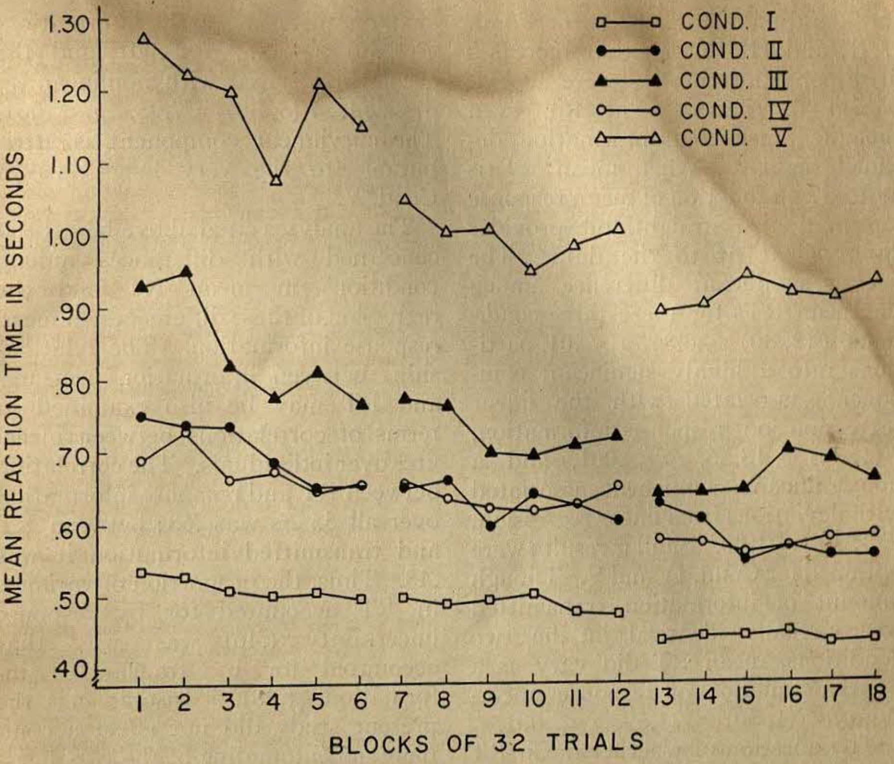


FIG. 1. Mean RT in seconds as a function of successive blocks of 32 trials of practice for the five experimental conditions.

positions of the groups, however, remained stable throughout the three experimental sessions.

The principal comparisons among the five conditions were based upon three measures of performance derived from the data of individual *S*s for the terminal six blocks of practice: (*a*) mean RT; (*b*) amount of information in the present response with information shared with the previous response partialed out—$H(R_2)$; (*c*) amount of information transmitted from the present stimulus to the present response with the information that both share with the previous response partialed out—$T_{R_1}(S_2; R_2)$. The particular information measures employed were designed to take into account first-order sequential dependencies in the response series generated by *S*s. Mean response measures for all conditions are summarized in Table 2.

Initially the results may be examined to compare RTs in conditions relatively alike with respect to transmitted information but differing considerably in response information. A

TABLE 2

MEAN RESPONSE MEASURES FOR THE EXPERIMENTAL CONDITIONS

Cond.	RT	$H(R_2)$	$T_{R_1}(S_2; R_2)$
I	.438	.999	.830
II	.573	1.798	.806
III	.662	2.546	.718
IV	.570	1.988	1.802
V	.920	2.490	1.767

consideration of the results for Cond. I, II, and III reveals that increases in response information were accompanied by increases in RT even though transmitted information declined slightly. When mean RT is plotted as a function of mean response information, a straight line provides an excellent fit to the data. The overall significant difference among the mean RTs for these three conditions, $F(2, 30) = 6.84, p < .01$, partitions into a highly significant component associated with the linear regression on response information, $F(1, 30) = 13.56, \ p < .001$, and a nonsignificant component associated with deviations from linear regression, $F(1, 30) < 1.00$. Similar results were obtained for Cond. IV and V. Though amount of information transmitted was virtually identical in the two conditions, mean RT did vary as a positive function of response uncertainty, $F(1, 30) = 25.59, p < .001$.

The question of whether mean RT can be described as a simple linear function of mean response information over *all* conditions must, however, be answered in the negative. In Cond. V mean RT is much longer than would be predicted from a regression line fit to the RT data of the first four conditions. A comparison between the data of Cond. III and Cond. V illustrates the same point. Despite near equivalence with respect to response information, the mean RT in Cond. III was .662 sec. as contrasted to .920 sec. for Cond. V. The *S*s who mapped four stimuli into eight responses reacted much more slowly than *S*s who mapped two stimuli into eight responses. The statistical significance of these results was established by partitioning the variance of mean RTs for the five conditions into two components, one associated with the linear regression on mean response information, $F(1, 30) = 45.25, \ p < .001$, and the second associated with nonlinear regression, $F(3, 30) = 7.95, \ p < .001$. The curvilinear component is attributable to the very long RTs in Cond. V.

The analyses cited above have been concerned with differences among conditions in mean RT and the regression of these differences on mean response information. The relationships between information measures and RT may be also examined in terms of correlations between measures over individuals. The correlation between RT and response information over all 35 *S*s was .68; between RT and transmitted information it was .48. Thus, the proportion of variance in RT accounted for by response uncertainty (.46) was twice that accounted for by transmitted information (.23). Insofar as the present study did not effect a complete unconfounding of the two measures—response and transmitted information correlated .32—it may be asked whether transmitted information made an independent contribution to the prediction of RT. The answer to this question is affirmative; the multiple correlation obtained when *both* information measures were used as predictors ($R = .73$) significantly exceeded the first-order correlation between RT and response information, $F(1, 32) = 5.52, \ p < .05$. Despite statistical significance, the increase in the proportion of the RT variance accounted for by the addition of transmitted information to the regression equation amounted to only .07, the difference between $.73^2$ and $.68^2$.

DISCUSSION

The information in a stochastic source is a weighted average of the number of binary decisions necessary to specify each possible message. This definition, coupled with the common empirical finding that RT is proportional to information measures, leads readily to the inference that information processing behavior involves a series of binary decisions with each decision taking unit time. Hick (1952) and Bricker (1955) have further suggested that the number of decisions made by *S* is better scaled by amount of information transmitted than by either stimulus or response uncertainty. Under this view, neither irrelevant stimulus information nor response equivocation should increase decision time. It is the "throughput" or information shared by the identifiable *external* stimulus and the response which is presumed crucial.

The results of the present experiment, however, demonstrate the existence of a task for which response information is a more adequate predictor of RT than is transmitted information. Why, in the present case, was RT more closely related to response uncertainty? The most obvious answer is that, in the response equivocation conditions, time was consumed by the process of random selection from among available correct response alternatives. Before this interpretation is considered at greater length, it must be noted that there is no logical necessity that randomization should require time. It might be argued, for example, that, were a random oscillation function to be superimposed on correct response alternatives of near-equal strength, reactions in an equivocation condition would be as rapid as those in a condition requiring no randomization. In other words, if an *S* were to respond with the first correct name to come to mind, and if this were not always the same name, information would be added to the response series without an increase in response latencies. This model of information processing behavior under conditions of response equivocation and the outcome predicted from it is consistent with the proposition that RT is proportional to amount of information transmitted; it is not consistent with the data obtained.

The finding that RTs were longer in equivocation conditions indicates that, in attempting to randomize response choices, *S*s do make decisions which require time. The *S*s must, in some sense, scan the list of correct response alternatives and/or a recent portion of the response series generated to a given symbol before arriving at a response choice. The decisions, then, are probably best conceptualized as responses to *internal* stimuli stored in memory. Since responses can share information with both external and internal stimuli, response uncertainty is potentially an excellent scalar of decision time in tasks in which the internal stimuli have implications for choice and are not simply noise. If it is supposed that decisions made to internal cues are like those made to external cues, both requiring the same order of time, then RT should be directly proportional to response information rather than to transmitted information. This result was obtained for the first four conditions of the present investigation; RTs, however, were much longer than might be expected in Cond. V.

In what way may the unpredictably long RTs in Cond. V be rationalized? In one respect Cond. V offered a special problem for *S*s. It was the only condition in which they were instructed to produce four random series simultaneously. If, under randomization instructions, selecting a response to an external stimulus involves the scanning of one or more previous responses to that stimulus, then Cond. V imposed a considerably greater tax on recent memory than did other conditions. With four possible symbols, the number of trials intervening between the occurrence of a stimulus and its recurrence would, on the average, be greater than that in a condition with only two symbols. When *S*s attempt to recall previous responses to a given stimulus, the retroactive effects of intervening responses to other stimuli may serve to increase the time required for the response selection process and consequently lengthen RT.

The results of the present investigation are not meant to stand in opposition to the proposition that RT is proportional to transmitted information. The decision is not whether this position is to be accepted or rejected, but rather to establish the limits of its applicability and, in so doing, to prepare the way for more comprehensive formulations. It is in this light, also, that recent research demonstrating independence of RT and transmitted information may be regarded.

REFERENCES

Archer, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

Bricker, P. D. Information measurement and reaction time. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955. Pp. 350–359.

Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Leonard, J. A. Tactual choice reactions. I. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

Merkel, J. Die zeitlichen Verhältnisse der Willensthatigkeit. *Phil. Stud.*, 1885, **2**, 73–127.

Morin, R. E., Forrin, B., & Archer, W. Information processing behavior: The role of irrelevant stimulus information. *J. exp. Psychol.*, 1961, **61**, 89–96.

Mowbray, G. H., & Rhoades, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **6**, 16–23.

Seibel, R. Discrimination reaction time as a function of the number of alternatives and of the particular stimulus-response patterns. *Amer. Psychologist*, 1959, **15**, 396. (Abstract)

(Received August 13, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 37–46

INFLUENCE OF UNBALANCED RESPONSE SCALES ON SIZE JUDGMENTS[1]

WALTER WEISS AND DAVID A. HODGSON

Boston University

Judgments of the relative sizes of 15 line-drawn squares were made by 20 and 72 college students in 2 experiments. Different groups used a 5-category response scale with 2 discriminating categories on each side of an interval designating "average" in size or with 3 discriminating categories on one or another side of this midpoint interval and a single nondiscriminating category on the other side. The results indicated a shift in the location of the "average" category of the scale relative to the series being judged, when the unbalanced scales were employed. Presenting an explicit standard for the midpoint category eliminated the main effects of the scales.

Studies of human judgment employing the method of single stimuli have customarily used response scales that have the same kinds and numbers of categories on both sides of the midpoint interval. For example, a 7-category scale of relative weight might consist of an explicit category for objects of average weight, flanked on each side by three comparably defined intervals for objects above or below average in weight. Even when the effects of different kinds of response scales are examined with this psychophysical method, the variants contain distributions of categories that are balanced around a midpoint (Hunt & Flannery, 1938; Nash, 1950; Weiss & Ten Eyck, 1960). What judgmental effects might occur if the scale has more categories on one side than another is unknown. And yet, in ordinary matters outside the laboratory, people often give evidence of more refined response discrimination among objects that are on one or another side of the midpoints of their scales. For example, a person may discriminate in words between people whom they consider above average in intelligence, calling a few "geniuses," others very bright, others quite intelligent, others merely average, and all the rest below average. With a little reflection, one can find other examples illustrating skew distributions of response categories. But, whatever the basis for the customary use of an unbalanced scale, what effects would the scale have *by itself* on judgments? Would effects be found with judgments of neutral, physical stimuli for which some objective estimate of the judged attribute is available? The first experiment to be reported was addressed to these questions and the second one to an elaboration of the findings.

EXPERIMENT I

Method

In individual sessions, 20 *S*s judged the relative sizes of a set of 15 line-drawn squares. Half of them used a 5-category response scale containing intervals labeled +3, +2, +1, 0, and − (to be referred to as the "unbalanced plus" scale); and half used categories designated +, 0, −1, −2, and −3 (to be referred to as the "unbalanced minus" scale). All were told with emphasis to use the "0" category for squares that were *average in size relative* to the rest of the series and to use the general plus or minus category for those above or below average in relative size, regardless of how much above or below

---

[1] The research reported in this paper was supported by Grant M1351 from the National Institute of Mental Health.

they might be. The three discriminating minus or plus categories were to be applied to squares that were "slightly," "quite," or "extremely" below or above the average in size.

To familiarize *S* with the range of square sizes to be judged and to accustom him to the apparatus, an orientation trial was provided during which he made no overt response. Although the entire set of 15 squares was presented in a randomized order on this trial, *S* was only told that he would see a set of stimulus cards of the kind he would be judging. At no time during the experiment was he informed by *E* that there were 15 different squares nor that there would be six judgment trials. Immediately after the orientation trial, the general instructions concerning the nature of the scale were repeated. Then the first of six judgment trials was begun. The order of the 15 squares varied from trial to trial according to predetermined random arrangements of presentation. The same set of 10 such orderings was used for each group; hence one *S* from each group had the same sequence of 90 stimuli presentations. A momentary delay occurred at the end of each trial to permit *E* to insert into the apparatus the prearranged and ordered set of 15 squares for the next trial. Immediately before the sixth trial, *E* casually mentioned that this would be the last trial. Hence, to *S* a trial meant the presentation of a considerable number of stimulus cards and a judgment of the relative size of the square on each.

After the last judgment, *S* was asked to estimate the total number of times he used each of the response categories on the last trial and the total number of different squares he had seen on this trial. Also he drew on three sheets of graph paper three squares: the first one represented his estimate of a square that was halfway between the smallest and largest seen; then, in counterbalanced order over the 10 *S*s in each group, two that represented the smallest and the largest seen.

TABLE 1

AVERAGE SIZE OF SQUARE PLACED IN ZERO CATEGORY ON EACH TRIAL (In. per Side): EXP. I

		Trials					
		1	2	3	4	5	6
Unbalanced plus	*M*	1.40	1.51	1.53	1.46	1.47	1.34
	SD	0.41	0.42	0.44	0.26	0.48	0.35
Unbalanced minus	*M*	1.74	1.72	1.76	1.82	1.87	1.97
	SD	0.51	0.43	0.47	0.41	0.46	0.54

TABLE 2

ANALYSIS OF VARIANCE OF AVERAGE SIZES OF SQUARES PLACED IN ZERO CATEGORY ON EACH TRIAL: EXP. I

Source	*df*	*MS*	*F*
Scales (S)	1	4.34	4.57*
*S*s/Groups (*S*s/G)	18	0.95	
Trials (T)	5	0.03	<1.00
T × S	5	0.11	2.75*
*S*s × T/G	90	0.04	
Total	119	0.12	

* $p < .05$.

The 15 squares ranged in extent of a side from .57 in. to 4 in., with the middle-sized one being 1.52 in. per side. They formed a geometric series in which the constant multiplier was $\sqrt[5]{2}$. They were presented one at a time at the wide end of a lighted tunnel having the approximate form of a frustum of a square pyramid. The *S* set his forehead against a restraining rest at the narrow end. In this position, his eyes were a few inches from a shutter that was raised when a stimulus was to be judged and dropped immediately after the judgment was given. Each stimulus was presented for a maximum of 2 sec. for judgment. The shutter was dropped at the end of that time if no response had been given, and *S* was then asked for an immediate judgment. The total interval between the onset of exposures for any successive pair of squares was 10 sec.

The 20 *S*s were students at Boston University who were paid for their time. They were assigned randomly to groups and sequences of stimuli presentations.

Results

Since the scales are asymmetrical in opposite directions, we are limited in the kinds of comparisons that can be made. Only the zero category is common to both. Hence, primary evidence for differential effects was sought in variations in the use of the "average" category. The median of the sizes of squares placed in that category was determined for each *S* for each trial. In Table 1 are the means of these medians for each group for each trial, and in Table 2 is an analysis of variance of these data.[2] The average size of square placed in the zero category by the group using the unbalanced plus scale is significantly smaller than that assigned by the other group (1.45 in. vs. 1.81 in.).[3] This difference appears on the first trial, diminishes slightly on the next two trials, and then increases steadily to its largest value on the sixth trial. In fact, on the last trial, when *S*s are matched according to order of presentation of the stimuli, 7 out of 10 pairs showed no overlap or crossing of the distributions of square sizes placed in the zero category.

These data suggest that the scales produced a divergence in the alignment of the average category of judgment relative to the series of stimuli being judged. If this is so, then the frequency of use of the combined plus and combined minus categories should be differentially affected. It would be expected that the total frequency of plus responses assigned by *S*s using the unbalanced plus scale would be greater than that assigned by the other group; the reverse would occur in regard to the total frequencies of minus responses. No expectation of difference need be held in regard to the frequency of use of the zero category. In Table 3 are the mean numbers of times these categories were used by each group over all trials. The analysis of variance of these data revealed a significant interaction *F* for scales by categories. By *t* tests, the groups are significantly different in their frequency of use of the plus category, almost so in their use of the minus category, and comparable in regard to the zero category. Since the average number of different squares placed in the zero category was not affected by the scales (3.2 vs. 3.5) and *S*s tended to place consecutively sized squares in the zero category, the data provide further support for the supposition that the results are due to a shifting of the zero interval relative to the stimulus series being judged.

TABLE 3

Frequency of Use of Plus, Zero, Minus Categories over All Trials: Exp. I

		Plus	Zero	Minus
Unbalanced plus	*M*	39.00	19.30	31.70
	SD	10.35	4.62	11.16
Unbalanced minus	*M*	28.00	21.40	40.60
	SD	13.05	6.02	9.54
p (two-tailed)		.05	>.30	<.10

Further confirming evidence on this major point is provided by the following analysis. On each of the last three trials, the difference was computed between the largest square placed by each *S* in the average category and the smallest in the first plus category above and, separately, the difference between the smallest square placed in the zero category and the largest in the first minus category below. The mean of the former three differences was taken as an estimate of the upper bound of *S*'s zero category; and the mean of the latter three, as an estimate of the lower bound. The means for the groups were: unbalanced plus, 1.16 and 1.80; and unbalanced minus, 1.57 and 2.44. The differences between the groups in the lower bounds and, separately, in the upper bounds of their zero category were significant beyond the .05 level.

To determine if the outcomes could be attributed to a natural tendency to use a rectangular distribution of assignments to the five categories (Parducci, Calfee, Marshall, & Davidson, 1960; Parducci & Marshall, 1961), chi square analyses were carried out based on expected frequencies of 18 per category. For each group, the sum of the chi squares for the *S*s was significant, $\chi^2(40) = 276.2$ and 236.2, *p*'s $< .001$, as was the group chi square, $\chi^2(4) = 164.88$ and 44.40, *p*'s $< .001$. The difference between these two types of chi square with $df = 36$ is designated by Snedecor (1956) as "heterogeneity" chi square and is a test of the consistency of the individual patterns of assignments. For each group, this chi square is significant at less than the .001 level providing evidence of interindividual differences. Examination of the seven individual chi squares in each group that were significant at or beyond the .05 level did not reveal an instance in which the pattern of assignments was the reverse of the general tendency to place too many stimuli in the nondiscriminating plus or minus category and too few in the aggregate of the discriminating categories, particularly in the ±2 and ±3 intervals. Individual differences arose in the extent to which the nondiscriminating categories or the zero category were overemphasized or the ±3 categories were underused. Hence, the data are not consistent with a hypothesis of a rectangular distribution of assignments.

---

[2] All *p* values reported in this paper are two-tailed.

[3] The square eighth in ascending order in the series was 1.52 in. per side and the square ninth in order was 1.74 in. Hence, these squares fall between the two averages.

It is possible however that a rectangular distribution was used with an *implicit* 7-category scale; i.e., *S*s may have held in mind a $+3$ to -3 scale even though they responded in terms of the given 5-category one. To check on this, chi square analyses were carried out with expected frequencies of 12.86 for the zero and the discriminating categories and 38.58 for the nondiscriminating one. The results again establish significant individual chi squares for 7 out of 10 of the *S*s in each group. The sums of the individual chi square (167.78 and 205.11) are significant at less than the .001 level, as were the group chi square (69.78 and 94.66). The primary tendency now is to underuse the nondiscriminating categories and to overuse the zero category, a pattern that accords with the supposition of a shift in the zero point of the judgment scale in the direction of the nondiscriminating category. Individual differences in the extent of this tendency and two reversals account for the significant heterogeneity chi squares that were found again.

The groups were comparable in the sizes of squares drawn as the smallest seen, the largest seen, and the one midway between the smallest and largest. The overall averages were: .75 in. per side, 3.51 in., and 1.99 in.; the actual sizes were: .57 in. per side, 4 in., and 1.52 in. There were also no differences between their estimates of the number of different squares seen on the last trial (an overall average of 5.6 was reported) or in the discrepancies between their actual use of the response categories on the last trial and their estimated frequency of use.

Discussion

The results of this study suggest that oppositely skewed distributions of response categories can produce divergent scale judgments of neutral, physical stimuli that are being judged on a physical dimension. Specifically, the groups differed in the locations of the midpoints of their judgment scale, the "average" category, relative to the series of stimuli being judged and in the direction of the side of the scale with the fewer categories. One possible explanation for the findings is that the *standard* adopted for the undifferentiated plus or minus category affected judgments of stimuli that might normally be placed somewhat above or below the average. That is, when a person places an extremely small square in the nondiscriminating minus category, he is less likely to assign the same nondiscriminating minus to a square that is much larger although still below the series average. The recalled difference is sufficiently great that some response differentiation is compelled. Similarly, the presence of smaller or larger than usual stimuli in the average category will cause some stimuli that might normally fall close to the opposite end of the "average" interval to exceed the latitude of response equivalence for this category and, hence, to be assigned to the next one. The availability of discriminating categories for stimuli that are above the series average prevents a directly comparable problem from arising in the assignments of the larger-than-average squares.

Since the stimuli at the ends of the series are readily and reliably discriminated from the others and are assumed to function as end anchors (Garner, 1952; Volkmann, 1951), they will be less affected by the changes in the scales than are the stimuli closer to the middle of the series. The data bear this out. For example, on the last trial, as square sizes moved in from the extremes of the very smallest or largest, the number of *S*s placing them in a relevant nondiscriminating category steadily decreased.

The proposed interpretation of the data emphasizes implicit standards and response factors as determinants of the results, rather than implicating perceptual changes. The comparability of the groups in their drawn estimates of the extreme and middle squares, in their judgments of the number of different squares seen on the last trial, and in their recall of the number of times they used each response category supports this assumption.

The data of this study do not reveal if both unbalanced scales affected the scale assignments of the squares in comparison with those that would be obtained with a balanced scale.[4] Furthermore, if the interpretation presented is more than plausible, we should expect little or no effect of unbalanced scales if an explicit standard is provided for the critical zero category. To obtain information on both of these points and to replicate the basic study, a second experiment was undertaken.

EXPERIMENT II

Method

The following aspects of method were the same as in Exp. I: the 15 squares, the basic procedure for presenting the stimuli, the postjudgment questions, and the drawings of estimates of certain square sizes. However, seven judgment trials were given instead of six.

Three types of scales were employed: the same two kinds of unbalanced scales as in Exp. I and a 5-category balanced scale ranging from $+2$ to -2. The ± 1 categories of the latter scale were to be used for squares that were slightly above or below average in size; and the ± 2 categories, for those that were quite above or below average in size; the zero category, of course, referred to those average in size relative to the series of stimuli being judged.

Half of the *S*s using each type of scale were presented with an explicit standard for the zero category, and half were not. The special instructions given the former *S*s were as follows:

> Now occasionally before a square to be judged is presented, I am going to show you a square which can be considered average in size relative to the other squares of the series you will be judging. I will tell you when this square is to be shown. If it were being judged, you would call it a "zero." I want you to remember that of course other squares in the series might also be judged to be generally average in size relative to the set and they would also be called "zero"; for people often place a range of sizes in the average category and not necessarily only one size.

On each trial the standard was presented for 2 sec. before the first, the sixth, and the

---

[4] In two previous studies (Weiss & Ten Eyck, 1960, 1961), in which balanced scales were employed with the same stimuli, the median stimulus placed in the midpoint interval was approximately 1.74 in. per side.

TABLE 4

Average Size of Square Placed in Zero Category: Exp. II

	Unbalanced Plus		Balanced		Unbalanced Minus	
	Standard	No Standard	Standard	No Standard	Standard	No Standard
M	1.58	1.71	1.59	1.77	1.61	2.19
SD	0.14	0.42	0.11	0.36	0.16	0.54

eleventh square. The standard was the same as the middle square of the series being judged.

The *S*s were 72 students taking courses during the 1961 summer session at Boston University; they were paid for their participation. They were assigned randomly as they came to one of the six groups and one of the 12 prearranged random sequences of the 105 stimulus cards.

Results

In Table 4 are the average sizes of squares placed in the zero category by each of the six groups. As might be expected (Volkmann, 1951), when an explicit midpoint standard is presented even from time to time, the variance of average square sizes placed in the zero category is considerably smaller than when a standard is not made available; the *F* ratio of variances has a $p < .001$. In general, the presence of the standard reduces the average square size placed in the midpoint category; by the sign test, based on the 36 matched pairs using the same scale, $p < .05$. Hence, the standard counteracts the tendency for adaptation level to be well above the midpoint of this stimulus series. By separate analyses of variance, the scales were found to have a significant effect only under the no standard condition (for *F* for scales $p < .05$). The average size of square assigned to the midpoint category by the unbalanced minus group was significantly larger than those assigned by each of the other two groups (*t* test $p < .05$), whose averages were comparable.

Under the no standard condition, the difference between the unbalanced minus group and each of the other two was significant on the first trial by the sign test, applied to pairs of *S*s matched by order of presentation ($p < .05$ for each comparison); the means were: unbalanced minus, 1.94 in.; balanced, 1.64 in.; unbalanced plus, 1.69 in. The differences decreased slightly on the next three trials and then increased over the last three to their largest values; on the last trial, the means were: unbalanced minus, 2.28; balanced, 1.85; and unbalanced plus, 1.60 in. On this last trial, the three no standard groups were significantly different from each other; both *t* tests and sign tests yielded $p < .05$ for all comparisons.

As in Exp. I, the lower and upper

TABLE 5

Upper and Lower Bounds of Zero Category: Exp. II

		Unbalanced Plus		Balanced		Unbalanced Minus	
		Lower	Upper	Lower	Upper	Lower	Upper
No standard	*M*	1.26	2.18	1.50	2.34	1.79	2.69
	SD	0.29	0.40	0.39	0.53	0.54	0.65
Standard	*M*	1.39	1.93	1.35	1.97	1.36	1.92
	SD	0.18	0.21	0.10	0.14	0.09	0.12

TABLE 6

Number of Squares Placed in Plus, Zero, Minus Categories on Each Trial: Exp. II

		Plus		Zero		Minus	
		M	*SD*	*M*	*SD*	*M*	*SD*
Unbalanced plus	Standard	5.80	0.86	2.87	1.39	6.33	1.20
	No standard	5.02	1.61	3.71	1.24	6.26	1.95
Balanced	Standard	5.99	0.70	2.46	0.94	6.55	0.69
	No standard	4.95	1.46	3.19	1.12	6.86	1.52
Unbalanced minus	Standard	6.06	0.92	2.21	0.97	6.73	0.70
	No standard	3.94	2.05	2.93	1.13	8.13	1.87

bounds of the zero category were determined for each *S* over the last three trials. The means are entered in Table 5. Under the standard condition, the scales had no effects on the limits of the zero category. But, under the no standard condition, the scales produced displacements of both bounds of the midpoint interval in the directions of the differences found for the comparisons based on the average sizes of squares placed in this interval. By analysis of variance, the *F* for groups was significant ($p < .05$) for the lower and the upper bounds, separately. By *t* tests, only the differences between the two unbalanced groups were significant (both p's $< .02$). But the regular progression of the limits as the number of plus categories decreased (or the number of minus increased) was confirmed by highly significant linear components of variance ($p < .02$) and insignificant quadratic ones.

Since the average size of square placed in the zero category and the bounds of this category were affected by the scales, we should expect differences in the average number of different squares placed above and below the midpoint category. The relevant data are in Table 6. The primary effect of the standard was to reduce the number placed in the zero category and to reduce the tendency to place more in the minus category than in the plus category. Analysis of variance indicated that, under the standard condition, the scales did not affect the distributions of assignments. But, under the no standard condition, they did; the scale by category interaction term is highly significant ($p < .01$). The unbalanced minus group overuses the minus category in an absolute manner and relative to the plus category to a significantly greater extent than does each of the other groups, which are not significantly different from each other; *t* test p's $< .05$. The no standard groups are not significantly different from each other in their frequency of use of the midpoint category. These results are consistent with the outcomes of the preceding analyses and corroborate the findings involving the unbalanced groups in Exp. I.

As in Exp. I, it seemed desirable to determine if a hypothesis of a rectangular distribution of assignments over the categories of the scale could account for the results. Chi square was computed for each *S* on the assumption of a theoretical value of 21 for each category. The sum of chi squares with $df = 48$ was significant

TABLE 7

Amount of Information Transmitted (I_t): Exp. II

		Unbalanced Plus	Balanced	Unbalanced Minus
Standard	*M*	1.72	1.85	1.67
	SD	0.11	0.08	0.10
No standard	*M*	1.46	1.73	1.67
	SD	0.11	0.13	0.20

for each group; all p's $< .01$. Under the no standard condition, 6/12 *S*s using the balanced scale had significant chi squares; 11/12 *S*s using the unbalanced plus scale had significant chi squares; and 8/12 *S*s using the unbalanced minus one had significant chi squares. The comparable proportions under the standard condition were 3/12, 12/12, and 12/12. All group chi squares with $df = 4$ were significant at less than the .001 level. Each of the no standard groups had a significant heterogeneity chi square ($p < .01$, $df = 44$); but for each of the standard groups, heterogeneity chi square was not significant. Hence, the presence of the standard reduced individual differences in the distribution of responses over the five categories of the scale. Under the no standard condition, individual differences occurred in the unbalanced groups in the extent to which the zero or nondiscriminating categories were overused relative to theoretical expectation. (A similar pattern occurred in Exp. I.) In the balanced group, individual differences occurred in the overuse of the minus categories and underuse of the +2 category. In sum, the data do not support an interpretation based on an assumed rectangular distribution of assignments.

To test an alternative hypothesis that, despite the required use of an unbalanced scale, the no standard *S*s used an implicit 7-category scale, chi square analyses were carried out. The theoretical frequencies were 45 in the nondiscriminating category and 15 in each of the others. A significant sum of individual chi squares was obtained for the unbalanced minus and separately, for the unbalanced plus groups (p's $< .001$). Group chi squares were also significant at less than the .001 level. Significant individual chi squares were exhibited by 9/12 *S*s using the unbalanced minus scale and 10/12 of those using the unbalanced plus scale. Both groups exhibited significant heterogeneity chi squares. But individual differences occurred primarily in the extent of underuse of the nondiscriminating category. The pattern of results is again quite like that obtained in Exp. I and impugns the validity of an interpretation based on the implicit use of a 7-category scale.

Following the procedure recommended by Garner and Hake (1951), an information analysis was carried out on the stimulus by response matrices of the *S*s. Analysis of variance of the data presented in Table 7 revealed highly significant *F*s for standard, scales, and the Standard × Scale interaction (all p's $< .005$). The *S*s using the standard had a significantly higher mean transmitted information (I_t) than did the no standard *S*s (1.74 vs. 1.62). This was due essentially to a significant difference in response equivocation (E_r) not in response information (I_r). The significant Standard × Scale interaction occurred because the presence of a standard increased I_t significantly when the balanced or unbalanced plus scale was used but had no effect when the unbalanced minus scale was used. In general, the use of a balanced scale led to a significantly higher I_t than the use of the unbalanced scales (t test p's $< .05$). Under the no standard condition, the unbalanced minus group had a significantly higher I_t than did the unbalanced plus. The latter two effects of the scales were due essentially to significant differences in I_r not in E_r.

Neither the scales nor the presence of a standard affected the sizes of squares drawn as midway between the smallest and largest seen, the smallest seen, or the largest seen. Analysis of variance yielded insignificant *F*s for all sources of variation except, of course, square size. The overall averages were: smallest, .61; middle, 1.88; and largest, 3.66 in. The actual sizes were: smallest, .57 in. per side; middle, 1.52 in.; and largest, 4 in.

Neither the scales nor the standard affected significantly *S*s' estimates of the number of squares of different sizes seen on the last trial. An overall average estimate of 6.20 was recorded. The estimated pattern of use of the categories on the last trial paralleled their actual frequency of use.

Discussion

The results of Exp. II confirm and extend those of Exp. I. Taken together, they demonstrate that scales with skewed distributions of response categories can produce divergent judgments of the same set of physical stimuli. Descriptively, the data suggest that a shift occurred in the location of the midpoint of the judgment scale relative to the stimulus series being judged and in the direction of the side of the scale with fewer categories. Although the unbalanced plus scale was not decisively different than the balanced scale in its effects, the increasing divergence on the last few trials, culminating in a significant final discrepancy, supports this fuller assertion. When an explicit standard is made available for the "average" category, thereby fixing it to a common portion of the physical continuum, the main biasing effects of the unbalanced scales are removed. The conceptual analysis set forth in the Discussion of Exp. I can be applied to the major outcomes of Exp. II in a relatively obvious way and therefore will not be repeated here.

The results of the information analysis can also be brought within the context of an interpretation based on the effects of major implicit and explicit category standards. As exemplification, the following two analyses are offered. Presenting an explicit standard reduces the number of alternative scale positions for a stimulus that is not extremely large or small; for the midpoint standard permits a more certain judgment concerning the side of the zero category to which the stimulus should be assigned or if it falls within the bounds of this category. Furthermore, its felt proximity to the standard for the midpoint and to the extreme squares which anchor the end categories narrows the range of intervals that can realistically be considered. This should occur each time the stimulus is presented for judgment. Hence, one general effect of an explicit midpoint standard should be to reduce response equivocation.

The contribution to I_r (response information) of a category of response decreases as the proportionate use of the category deviates from .20 when a 5-category scale is used. Also, more than one fifth of the stimuli arranged in a geometric series would ordinarily be judged above average relative to the rest of the series. On the basis of the general explanation of the effects of the scales, the unbalanced minus scale when employed without the explicit middle standard would be expected to reduce the number of different squares placed into the above average category below what would ordinarily be obtained. However, the explicit standard by fixing the center of the average category and narrowing its range restores the more normal, adverse (relative to .20) overuse of the plus category. Hence, I_r should be higher without the standard than with it, as did occur. And, since E_r

(response equivocation) is less when the standard is available, the net result should be a relatively unchanged I_t for the unbalanced minus scale. That a comparable result did not occur when the unbalanced plus scale was used is attributed to the greater difficulty of discriminating between the smaller sized squares. This maintained the more customary overuse of the below average category; and therefore, I_r was not affected by the standard.

To be sure, the explanations offered are post hoc in regard to the data of these experiments. But, their value lies in a common, unifying context of emphasis on the effects of category standards and the structure of the response system. This type of analysis leads readily to predictions about the effects of various kinds of scales and distributions of stimuli on other types of judgments.

If adaptation level is construed as the stimulus at the midpoint of the judgment scale or the stimulus judged "average" (Helson, 1947, 1948, 1959), then the data reported do not accord with a direct application of AL theory. For, without a change in the stimulus series presented for judgment or in background and context factors or in residual factors, AL was experimentally modified. The structure of the response scale was the manipulated variable. Hence, as others have found (Eriksen & Hake, 1957; Volkmann & Engen, 1961), response factors are basic to the judging process; and any general theory of judgment will have to give them more than passing reference.

REFERENCES

ERIKSEN, G. W., & HAKE, H. W. Anchor effects in absolute judgments. *J. exp. Psychol.*, 1957, **53**, 132–138.

GARNER, W. R. An equal discriminability scale for loudness judgments. *J. exp. Psychol.*, 1952, **43**, 232–238.

GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.

HELSON, H. Adaptation-level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A Study of a science.* Vol. I. New York: McGraw-Hill, 1959. Pp. 565–621.

HUNT, W. A., & FLANNERY, J. Variability in the affective judgment. *Amer. J. Psychol.*, 1938, **51**, 507–513.

NASH, M. C. An experimental test of the Michels-Helson theory of judgment. *Amer. J. Psychol.*, 1950, **63**, 214–220.

PARDUCCI, A., CALFEE, R. C., MARSHALL, L. M., & DAVIDSON, L. P. Context effects in judgment: Adaptation level as a function of the mean, midpoint, and median of the stimuli. *J. exp. Psychol.*, 1960, **60**, 65–77.

PARDUCCI, A., & MARSHALL, L. M. Context-effects in judgments of length. *Amer. J. Psychol.*, 1961, **74**, 576–583.

SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Iowa State Univer. Press, 1956.

VOLKMANN, F. C., & ENGEN, T. Three types of anchoring effects in the absolute judgment of hue. *J. exp. Psychol.*, 1961, **61**, 7–17.

VOLKMANN, J. Scales of judgment and their implication for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 273–294.

WEISS, W., & TEN EYCK, P. H. Effects on judgments of changes in the response scale. *Psychol. Rep.*, 1960, **7**, 487–495.

WEISS, W., & TEN EYCK, P. H. Effect of shock on size judgments. *Psychol. Rep.*, 1961, **9**, 135–144.

(Received August 15, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 47–52

CONDITIONING OF MOTOR AND VERBAL RESPONSES TO NONVERBAL STIMULI[1]

W. A. BOUSFIELD AND T. M. COWAN

University of Connecticut

College student Ss individually underwent one of 2 methods of training. One involved actual conditioning of a hand withdrawal response. The UCS was a falling object, and the CS a flash of light. Alternatively, the Ss merely observed the sequential occurrence of CS and UCS. Both types of training were effective in establishing connections between verbal associates elicited, respectively, by CS and UCS. Evidence of the effect was obtained by means of forced-choice verbal association tests. In each test item S was given an associate of the CS (or UCS). He then made his choice of a related word from 5 words, 1 of which was an associate of UCS (or CS).

This is a report of a study designed for testing two experimental hypotheses: (*a*) The sequential occurrence of a nonverbal CS and a nonverbal UCS according to the requirements for the classical conditioning of a motor response should result not only in the conditioning of the motor response, but also in the establishing of connections between the verbal associative responses elicited, respectively, by the CS and the UCS. (*b*) When *S*s are instructed merely to observe the sequential occurrence of two nonverbal stimuli, the consequences should include the establishment of connections between the verbal associative responses elicited, respectively, by the two nonverbal stimuli. The procedure for testing both these hypotheses made use of an apparatus in which the nonverbal stimuli were presented. The CS was a flash of light. The UCS was a golf ball observed by *S* to be falling from the top to the bottom of the apparatus. When motor conditioning took place *S* placed his outstretched hand in the apparatus so that he would be struck by the falling ball unless he withdrew his hand rapidly. Use was made of the following verbal associates of the CS: LIGHT, FLASH, and BLINK. The associates of the UCS were DROP, BALL, and FALL.

Both the experimental hypotheses with which the present study is concerned can be derived from a combination of experimental evidence and theory. In the first place, as was reported by Karwoski, Gramlich, and Arnot (1944), nonverbal stimuli such as objects and pictures of objects are capable of eliciting verbal associational responses similar in kind to those which are elicited by words. Two extensions of this finding are warranted. It appears reasonable to assume that any nonverbal stimulus capable of being discriminated and labeled is associatively connected with verbal responses. Furthermore, it may be assumed that such verbal responses are implicitly elicited or primed automatically and immediately. The second of these assumptions is in effect an extension to nonverbal stimuli of the postulates proposed by Bousfield, Whitmarsh, and Danick (1958) in their account of responses to words. Proceeding further in the derivation of the experimental hypotheses, it may be inferred that when nonverbal stimuli occur sequentially, the implicit verbal responses they elicit must also occur sequentially. This is a sufficient condition for the establishment of linkages between the associative responses elicited by the respective sequentially occurring nonverbal stimuli. If such an interpretation is valid, the number of associative linkages established in this way must be large indeed. It is likely that most of the connections established in this way are very weak. This is no argument against their existence though it does follow that sensitive devices are needed for their detection.

---

[1] This is Technical Report No. 38 under Contract Nonr-631(00) between the Office of Naval Research and the University of Connecticut. Reproduction in whole or in part is permitted for any purpose of the United States Government.

No mention has so far been made of the conditioning of a motor response which the first of the two experimental hypotheses indicates should occur simultaneously with verbal conditioning. As long as the requirements for classical conditioning are met, the attachment of a motor response to a new stimulus must of necessity take place and a theoretical account of the learning appears unnecessary in the present discussion. As for two types of learning taking place at the same time, it may be submitted that there appears to be no reason why this should not take place as long as the responses involved do not conflict.

METHOD

Subjects.—The *S*s were 120 undergraduate students divided into six groups of 20 each. An additional group of 86 *S*s gave the verbal associative responses needed in the testing phase of the experiment.

Apparatus.—The apparatus for presenting the stimuli was a plywood box with one open side through which *S* could place his hand, palm down, over a hole 2½ in. in diameter located in the center of the floor of the box. A golf ball with an embedded screw was held 19 in. above the hole by means of an electromagnet. When the circuit through the magnet was broken, the golf ball dropped down and through the opening in the floor of the apparatus. Two flashlight bulbs were mounted adjacent to the electromagnet so they could not be seen by *S*. By operating a single switch *E* could produce a visible flash of light in the apparatus and ½ sec. later, while the lights were still on, release the golf ball.

Verbal items.—Two sets of verbal items were needed for the testing phase of the experiment. These were the verbal associates to the respective nonverbal CSs and UCSs. In collecting both sets of data *E* instructed the 86 *S*s as follows. They were told *E* would present a series of stimuli some of which were words while others were nonverbal. The latter would be things that would happen in the experimental apparatus which was placed on a table in the front of the room. As *E* presented each stimulus, the *S*s were to write the first word it made them think of. For preliminary practice *E* spoke the word TABLE, and then sounded a cricket clicker. He stated that illustrative responses to TABLE were chair and food.[2] Responses to the clicker might be the word sound or the word click. Having attempted to insure understanding of the instructions *E* then presented in serial order the following spoken words and nonverbal stimuli: MAN, RIVER, dropping of the golf ball in the apparatus, BOOK, HORSE, and flashing of the lights in the apparatus. The tabulations of the data indicated that the most common responses to the falling golf ball and their cultural frequencies given as percentages of 86 were: drop, 25.6; ball, 17.4; fall, 11.6. The corresponding data for responses to the flashing light were: light, 55.8; stage, 5.8; flash, 4.7; blink, 4.7. The unusual response "stage" was apparently attributable to the context effect of the apparatus on the platform occupied by *E*. For the purposes of the experiment, stage was discarded leaving light, flash, and blink as the primary verbal responses to the flashing of the light.

An additional two lists of 12 verbal items each were chosen as control words for the testing phase of the experiment. The first list was for the three groups given the test for forward verbal association, and the second for the three groups tested for backward association. Since there were three presumably interrelated associates for the CS and three for the UCS, the control items comprised subgroups of three interrelated words. They were chosen on the basis of the judgment of the two *E*s that they had no obvious associative relationship to the verbal responses to the CS or to the UCS. Furthermore, the subgroups of items were essentially unrelated to each other. The words used with LIGHT, FLASH, and BLINK were: (*a*) thief, doctor, soldier; (*b*) ocean, mountain, river; (*c*) carpet, chair, table; (*d*) whistle, quiet, music. The words used with DROP, BALL, and FALL were: (*e*) eagle, lion, sheep; (*f*) joy, afraid, anger; (*g*) carpet, chair, table; (*h*) whistle, quiet, music. The last two subgroups are repeated in the two lists. No. *S*, however, was given both lists.

---

[2] Words treated as stimuli are presented in small capital letters and words treated as responses are in lower case letters.

Procedure.—The six groups of *S*s were designated as follows: (I) motor response conditioning with test for forward verbal association; (II) motor response conditioning with test for backward verbal association; (III) observing with test for forward verbal association; (IV) observing with test for backward verbal association; (V) control with test for forward verbal association; (VI) control with test for backward verbal association. The *S*s in Groups I and II received similar instructions for the motor conditioning. Each *S* was told to place his right hand palm down so as to cover the hole in the floor of the apparatus. It was indicated that several occurrences would take place in the apparatus and that *S* was to make what would be regarded as a normal response. Care was taken to avoid the use of any word associated with the stimuli employed in the experiment. The *E* presented the CS (the flash of light) alone twice. The UCS (golf ball falling toward *S*'s outstretched hand) was then given. If *S* did not withdraw his hand to avoid being hit, *E* demonstrated by releasing the golf ball and quickly pulling *S*'s hand out of the way. The UCS was once more given to insure that *S* made the withdrawing response immediately and without the silent prompting. The *S* was then informed he had made the normal response and he was expected to continue to do so. Five training trials were given. On the sixth trial the CS was presented alone to test for the establishment of conditioning. All *S*s conditioned. Individual differences in the speed of conditioning were not considered.

The second phase of the experiment immediately followed the first. The *E* told *S* he would be given a short test of word association. This involved the use of a 9-page booklet. In the case of the testing for forward verbal association the word LIGHT, a CS associate, appeared at the top of 3 of the pages. Similarly, there were 3 pages for FLASH and 3 for BLINK. Directly below each of the CS associates were five additional words arranged in a column. One of these, the embedded response word, was one of the three UCS associates, namely, drop, blink, or fall. The remaining four items were selected at random from the appropriate list of control words. The *S* was told to look at the word at the top of the page, and then to select, from the five words below it, the one the top word most made him think of at the present time. The *S* was to indicate his choice by writing an X in front of the chosen word. The choice of the embedded response word was scored as correct. The nine combinations of stimulus words and embedded response associates on the 9 pages of the test booklet were as follows: LIGHT-drop, LIGHT-ball, LIGHT-fall, FLASH-drop, FLASH-ball, FLASH-fall, BLINK-drop, BLINK-ball, and BLINK-fall. The pages of the test booklets were separately randomized for each *S*. The procedure for testing for the presence of backward verbal association was similar to that which was used for the forward condition except for the reversal in the test booklets of the stimulus words and the embedded response associates.

The *S*s in Groups III and IV merely observed the sequential occurrence of the nonverbal CS and UCS. No attempt was made to obtain conditioning of the hand withdrawal. The *E* told each *S* to give close attention to the occurrences in the apparatus. The CS (the flash of the light) was presented twice. There was a pause of about 10 sec. after which *E* presented the UCS (the golf ball was dropped). This was followed by five CS-UCS sequences, and finally there was a single presentation of the CS alone.

The *S*s in Control Groups V and VI were members of a large class and all took part in the experiment at the same time. Each *S* received either a forward test or a backward test booklet without any knowledge of the procedure used with other groups. The instructions for the forced-choice associations were the same as for the experimental groups.

RESULTS

The data needed for analysis were contained in the test booklets which had been given to all six groups of *S*s. The measure chosen for statistical

TABLE 1

Results of Forced-Choice Tests of Forward Verbal Association

Cond.	Proportion of *S*s Selecting UCS Associate									Mean No. UCS Associates Selected per *S*
	LIGHT drop	LIGHT ball	LIGHT fall	FLASH drop	FLASH ball	FLASH fall	BLINK drop	BLINK ball	BLINK fall	
I Motor conditioning	.65	.50	.50	.75	.45	.65	.60	.60	.80	5.50
III Observing	.50	.40	.50	.50	.40	.50	.70	.65	.85	5.00
V Control	.30	.05	.30	.40	.20	.40	.70	.20	.45	3.00
I − V	.35	.45	.20	.35	.25	.25	−.10	.40	.35	2.50
III − V	.20	.35	.20	.10	.20	.10	.00	.45	.40	2.00

Note.—Stimulus words in small capital letters are CS associates; response words in lower case letters are embedded UCS associates.

analysis was the mean number of correct responses checked by a given group of *S*s. Tables 1 and 2 present the data in terms of the proportions of *S*s in each group who checked the correct response for any given word pair. Thus, 13 of the 20 *S*s in Group I who took the test for forward verbal association chose "drop" as the appropriate response to LIGHT. As Table 1 indicates, the proportion selecting the correct response was .65. For the control *S*s in Group V the proportion choosing the same correct response was .30. The mean number of correct responses checked per *S* also appear in Tables 1 and 2.

The first step in the treatment of the results was to apply an analysis of variance to the data obtained from the tests for forward association given to Conditioned Group I, Observing Group III, and Control Group V. The three groups were found to differ significantly, $F(2, 57) = 4.586, p < .05$. The means of the number of correct responses for Conditioned Group I and Observing Group III were significantly higher than the mean for Control Group V, respectively: $t\ (38) = 3.141, p < .01$; and $t(38) = 2.401$, $p < .05$. The difference between Group I and Group III, however, was not significant. The data were similarly analyzed to determine the extent to which backward associations had been established. An analysis of variance was also made of the data

TABLE 2

Results of Forced-Choice Tests of Backward Verbal Association

Cond.	Proportion of *S*s Selecting CS Associate									Mean No. CS Associates Selected per *S*
	DROP light	BALL light	FALL light	DROP flash	BALL flash	FALL flash	DROP blink	BALL blink	FALL blink	
II Motor conditioning	.45	.55	.40	.65	.55	.50	.30	.50	.25	4.15
IV Observing	.55	.85	.70	.75	.65	.75	.65	.70	.80	6.40
VI Control	.20	.35	.25	.30	.40	.20	.25	.25	.20	2.40
II − VI	.25	.20	.15	.35	.15	.30	.05	.25	.05	1.75
IV − VI	.35	.50	.45	.45	.25	.55	.40	.45	.60	4.00

Note.—Stimulus words in small capital letters are UCS associates; response words in lower case letters are embedded CS associates.

plied by Conditioned Group II, Observing Group IV, and Control Group VI. The differences between these groups were significant, $F(2, 57) = 15.5$, $p < .001$. The use of t tests showed that both Group II and Group IV made significantly more correct choices than the controls in Group VI, respectively: $t(38) = 2.41$, $p < .05$; and $t(38) = 6.250$, $p < .001$. Observing Group IV showed significantly more backward association than Conditioned Group II, $t(38) = 2.845$, $p < .02$.

The question arises of whether there were significant differences between the forward and the backward verbal associational effects. Direct comparisons of Groups I and II (motor conditioning) and of Groups III and IV (observing) did not appear feasible in spite of the absence of a significant difference between the control groups V and VI. The obtained mean score of 2.40 for Group VI did not differ significantly from the expected mean of 1.80 for the forced-choice tests. The difference, however, between the obtained mean of 3.00 and expected mean of 1.80 for Group V was significant, $t(19) = 2.797$, $p < .02$. What were termed difference scores were used for this analysis. In the case of the forward associational effects induced in Conditioned Group I and Observing Group III the method was as follows. The mean number of correct choices on each of the 9 pages of the test booklets for forward association given to Group V was subtracted from the corresponding number of correct choices of each individual *S* in Groups I and III. Similarly, difference scores were also computed for *S*s who took the tests for the backward associational effects. The mean scores of Control Group VI were subtracted from the scores of the individual *S*s in Conditioned Group II and Observing Group IV. The difference scores for each *S* were then arranged in four distributions for Groups I, II, III, and IV, respectively. An analysis of variance indicated no significant differences between the results of the tests for forward and for backward association. On the other hand there was a significant interaction, $F(1, 76) = 4.771$, $p < .05$. Consequently, t tests were applied to assess the significance of the differences between the means for all six of the possible comparisons for the four groups of *S*s. Table 3 shows the results of these tests. As may be seen, Observing Group IV was responsible for the interaction effect. The most potent effect found in the experiment was thus that of the formation of backward verbal associations in *S*s who merely observed the sequential occurrence of the nonverbal CS and UCS.

TABLE 3

COMPARISONS OF MEANS OF DIFFERENCE SCORES FOR EXPERIMENTAL GROUPS

Group Comparison	Difference	$t(38)$
IV > II	2.25	2.845**
IV > III	2.00	2.286*
IV > I	1.50	1.870
I > II	.75	.845
I > III	.50	.519
III > II	.25	.266

Note.—Group I = motor conditioning with forward test, Group II = motor conditioning with backward test, Group III = observing with forward test, Group IV = observing with backward test.
* $p < .05$.
** $p < .01$.

DISCUSSION

The results of the present study support the experimental hypotheses. The conditioning of a motor response by using a nonverbal CS and a nonverbal UCS resulted in the predicted establishment of parallel associative connections between members of the family of verbal associates to the CS and members of the family of verbal associates to the UCS. Furthermore, as was predicted, observation alone of the sequential occurrence of the nonverbal CS and UCS used for the motor conditioning was also effective in establishing associative linkages between the families of verbal associates elicited, respectively, by the CS and the UCS. Contrary to what was anticipated was the lack of a significant overall difference between the forward and the backward verbal associational effects. Furthermore, the data revealed an unexpected interaction. The *S*s who merely observed the CS-UCS sequence and were given the tests for backward association chose the largest number of correct responses. No ready explanation is available for these unexpected findings.

An important problem arising from the present study is that of the status of what may be termed the *labeling response*. The words light and flash, though given after instructions for free association to the nonverbal CS, the flashing of the light, were labels for what the *S*s observed. The same is also true of drop, fall, and ball which were recorded as free associates of the nonverbal UCS, the dropping of the ball. Such responses as blink, dark, and candle to the CS, and game, throw, and catch to the UCS, on the other hand, appear to be more clearly associational in nature. It may be said that when *S*s are instructed to give free associative responses to printed words the labeling response of the word itself is typically not emitted or produced overtly. It is of course possible, when obtaining free associational responses to nonverbal stimuli, to tell the *S*s they should avoid giving the name of the stimulus. According to the *partial response identities* theory, which was recently summarized by Bousfield (1961), the immediate response to a word is its implicit emission. The present interpretation is that this is actually a labeling response. Such a response is representational in nature and it produces feedback, thus providing the representational stimulus. Associational responses are attached to the representational stimulus by conditioning. This interpretation makes possible the useful assumption that a word functions both as a response and as a stimulus. The representational response develops as a consequence of repeated constant stimulation, and it is typically a fractional part of the total response to a given stimulus. These properties provide a basis for the assumption that verbal representational processes are regularly involved in various mental activities.

The partial response identities theory may be extended so as to account for verbal responses to nonverbal stimuli. It must be supposed that the familiar nonverbal stimulus regularly elicits a nonverbal representational process. In addition, however, it may be assumed that the familiar nonverbal stimulus should elicit one or more verbal representational responses in the form of labels for the given stimulus. Such verbal and nonverbal representational processes serve to represent a stimulus when it is not physically present. The more a given label has accompanied a given nonverbal stimulus the greater is its capacity to activate its nonverbal counterpart representational response, and to elicit a family of related associative responses.

REFERENCES

BOUSFIELD, W. A. The problem of meaning in verbal learning. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 81–109.

BOUSFIELD, W. A., WHITMARSH, G. A., & DANICK, J. J. Partial response identities in verbal generalization. *Psychol. Rep.*, 1958, **8**, 703–713.

KARWOSKI, T. F., GRAMLICH, F. W., & ARNOT, P. Psychological studies in semantics: I. Free association reactions to words, drawings and objects. *J. soc. Psychol.*, 1944, **20**, 233–247.

(Received August 16, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 53–56

SECONDARY REINFORCEMENT EFFECTS AS A FUNCTION OF METHOD OF TESTING[1]

WILLIAM F. REYNOLDS
Queens College, Flushing, N. Y.
JOYCE E. ANDERSON
Western Michigan University
AND NORMA F. BESCH
Columbus Psychiatric Institute, Ohio State University

2 experiments were performed to test the notion that discriminative stimulus training is a necessary condition for the establishment of a neutral stimulus as a secondary reinforcer. Experiment I employed 25 Y maze test trials and showed that single-stimulus training was sufficient to establish a secondary reinforcer provided that sufficient test trials were run to allow competitive effects to dissipate. Experiment II employed T maze test trials and showed maximum test performance for both single-stimulus and discriminative-stimulus groups under conditions in which maximum control was exerted over extraneous test effects.

A recent experiment by McGuigan and Crockett (1958) found that a stimulus, in order to serve as a secondary reinforcer, must first be established as a discriminative stimulus. These investigators gave discrimination training to one half of the *S*s and single-stimulus training to the other one half. During 10 test trials in a Y maze, the only evidence of secondary reinforcement effects was obtained with the discrimination trained *S*s. Experiment I of the present report is a replication of the McGuigan and Crockett study except for the fact that 25 test trials, rather than 10, were administered during the Y maze test.

EXPERIMENT I

Method

Subjects.—The *S*s were 56 female albino rats donated by the Upjohn Drug Company; they were between 90 and 110 days of age at the start of the experiment.

Apparatus.—The training apparatus was adapted from the discrimination trainer reported by Grice (1948). The *S* proceeded from a start box through an alley into a triangular shaped choice chamber which terminated in two goal boxes. The goal-box colors were black and white, and the entrance to each box was covered by a curtain of the same brightness as the particular goal box. The rest of the apparatus was painted medium gray, except for the floor sections of the choice chamber which matched the colors of the abutting goal boxes. The incentive, contained in a glass coaster, was a .1-gm. pellet of wet Purina mash. The test apparatus was a simple Y maze with gray start box, alley, and 10-in. arms which terminated in black or white, curtained, goal boxes.

Procedure.—Following handling and the establishment of a 23-hr. food privation schedule, *S*s were randomly assigned to the four experimental groups, 14 *S*s per group. Secondary Reinforcement Training (SRT) *S*s were reward trained for 75 trials by having both baited goal boxes of the same brightness for a given *S*. For SRT-I *S*s, the empty glass coaster was placed in the training goal box during the Y maze test trials, and for SRT-II *S*s the food cup was removed.

Discriminative Stimulus Training (DST) *S*s received 75 rewarded trials to either the black or white goal box, and 25 forced, non-

---

[1] Portions of this paper were presented at the Eastern Psychological Association convention, 1959. This research was supported, in part, by Grant NSF G-9793 from the National Science Foundation.

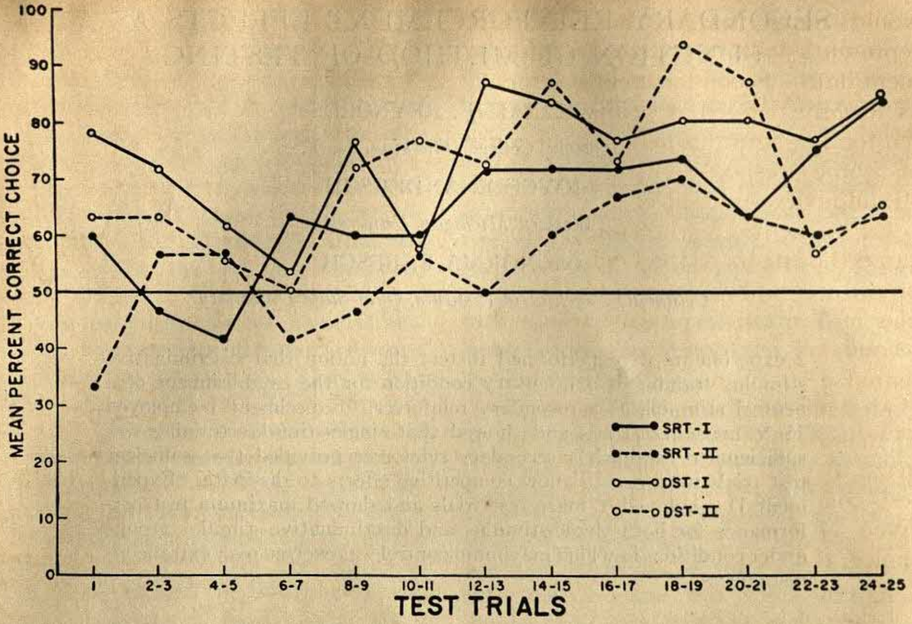


FIG. 1. Performance curves for the four experimental groups showing mean percentage choice of the positive stimulus during 25 Y maze test trials.

rewarded trials to the opposite color box. During the test trials, the empty food cup was placed in the positive box for DST-I *S*s, but was removed for DST-II *S*s. In each of the four experimental groups, an equal number of *S*s were trained to each goal-box brightness. The nonrewarded test series began on the eighth day following 5 reward training trials, and consisted of 25 Y maze trials with a minimum intertrial interval of 5 min. The positions of the positive and negative boxes in the maze were randomly determined.

Results

The results are summarized in Fig. 1 where it may be seen that group performances during Trials 1–10 differ from group performances during Trials 11–25. Statistical analyses of the data, by means of simple t tests and analysis of variance, revealed that (*a*) during Trials 1–10 both DST groups, but neither SRT group, performed significantly better than chance at the .05 level; (*b*) during Trials 11–25 all experimental groups performed significantly better than chance at the .05 level; (*c*) during Trials 1–25 only the SRT-II *S*s failed to differ significantly from chance performance, $t(13) = 2.10$, $p > .05$; and (*d*) for the entire test series there was a significant main effect of training condition in favor of the DST procedure, $F(1, 52) = 8.60$, $p < .01$, but the main effect of the food cup variable was not significant, $F(1, 52) = 1.97$, $p > .10$, and the interaction F was less than 1.00.

The results of Exp. I confirm the findings of McGuigan and Crockett, but in addition indicate that single-stimulus training can also produce secondary reinforcement effects when sufficient test trials are allowed for the dissipation of the competitive effects of factors like stimulus novelty (Kendler, 1959, p. 75). Since the normal use of the Y maze permits S to attend to the alternative stimuli while making a choice, many psychologists would deny that a Y maze test represents valid secondary reinforcement data. Secondary reinforcement, as usually defined, specifies that the reinforcing stimulus be a consequence of some response rather than a stimulus determinant of it.

Experiment II, therefore, investigated the effects of the DST and SRT procedures under each of two schedules of T maze test trials, where the secondary reinforcer was only presented *after* *S* completed the choice. Under one condition, all *S*s were given 15 massed, free-choice trials in the T maze (Cond. F); under the second test condition 4 trials per day were given, 1 free-choice and 3 forced-choice, for 15 days (Cond. C).

EXPERIMENT II

Method

Subjects.—The *S*s, who averaged approximately 120 days of age at the start of the experiment, were 40 male albino rats from the Rutgers University colony.

Apparatus.—The training apparatus was an exact replica of the discrimination trainer employed in Exp. I. The white and black goal boxes used in the training and test series, the incentive amount during training, and the food privation schedule were also identical with Exp. I conditions. The test apparatus was a wooden T maze, medium gray throughout, with start box, 18-in. alley, and 12-in. arms. Narrow spaces were provided on both sides of the choice point and at the terminal ends of the maze arms for the side insertion of gray plastic doors.

Procedure.—The DST and SRT training schedules were the same as in Exp. I except for two changes. First, only 65 reward trials were given, with 10 per day for 6 days and 5 additional trials on the first test day. Secondly, DST *S*s were not forced to enter the negative goal box a given number of times; no errors were made in the DST group after Trial 40. Ten *S*s from each training condition were run under each of two test trial procedures. Condition F *S*s were given 15 massed free-choice trials following the final 5 reward trials (DST-F, SRT-F). Condition C *S*s received 4 trials per day for 15 days, with Trial 1 a free-choice trial and Trials 2–4 forced-choice (DST-C, SRT-C). The forced-choice trials, administered in varying orders, insured that each *S* made two choices per day to each maze arm. All test trials were unrewarded but the empty food coaster was placed in the training goal box.

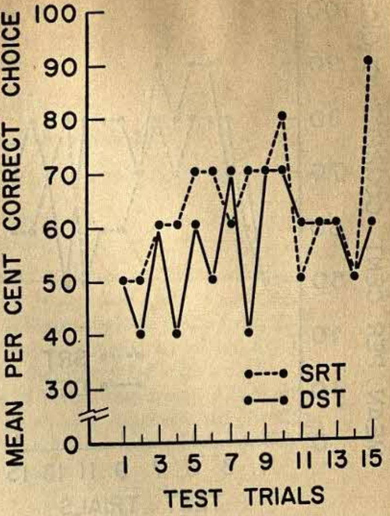


FIG. 2. Performance curves for the two groups tested under Cond. F (15 massed free-choice trials in the T maze).

One half of the *S*s in each group were reward trained to the white goal box, and one half to the black goal box. On all free-choice test trials the gray doors were in place before the goal boxes so that *S* could not see the goal-box color until after the choice response. Once *S* had made a choice, the choice point side door was inserted behind him and the goal-box door was then removed. Forced-choice trials for Cond. C *S*s were accomplished by blocking one maze arm with the appropriate choice point door. On Free Trial 1, to control for position preferences and to insure that the choice curves would begin at 50% or chance level, one half of the *S*s in each of the four groups found the training reward box on the side of choice, and one half of the *S*s found the opposite brightness box on the first turn side.

Results and Discussion

Performance curves, with percentage of choice of the positive training stimulus plotted against free-choice trials, for the SRT-F and DST-F groups are presented in Fig. 2. An inspection of the figure reveals that

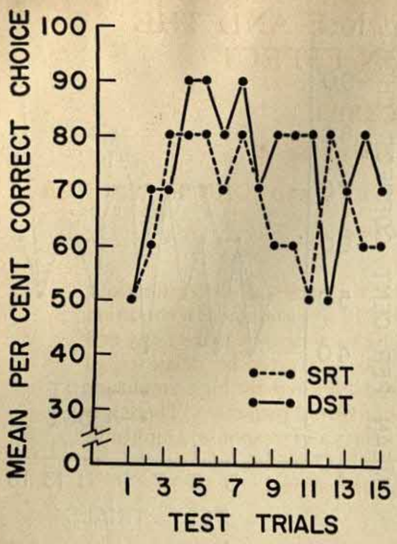


FIG. 3. Performance curves for the two groups tested under Cond. C (15 controlled free-choice trials in the T maze).

the SRT-F *S*s performed at a generally higher level than the DST-F *S*s. This unusual result is partly attributable to the fact that DST *S*s did not receive systematic avoidance training to the negative cue as they had in Exp. I. The mean number of correct choices for the DST-F group was not significantly higher than the 7.5 choices expected by chance, $t(9) = 1.50, p > .10$, but the mean of the SRT-F group exceeded chance at the .05 level, $t(9) = 2.53, p < .05$.

Performance curves for the SRT-C and DST-C groups are shown in Fig. 3 where it may be seen that both groups were, for the most part, consistently above 50% choice. The mean numbers of correct choices for both groups were significantly greater than chance at better than the .01 level. A simple factorial analysis of variance was also performed on the choice data and only the main effect of test trial procedure was found to be significant, $F(1, 36) = 5.63, p < .05$; the F for training condition was less than 1.00, and the Training Condition × Test Condition interaction was not significant, $F(1, 36) = 2.62, p > .10$.

The results of Exp. II indicate, in general, that secondary reinforcement effects (greater than chance performance) in maze learning can be produced both by single-stimulus training (Hull, 1943), and by discrimination training (Keller & Schoenfeld, 1950). The present results also indicate, however, that conclusions concerning the effectiveness of various training conditions in establishing secondary reinforcers must take account of the test methods employed in assessing such effects. The superiority of the Cond. C test method in Exp. II is partly due to the control of the attractive or avoidant effects of the novel alternative stimulus on the SRT *S*s. In addition, the Cond. C technique also provides some control of avoidance responses induced by frustration of the training box r_g-s_g during nonrewarded test trials for SRT *S*s (Amsel, 1958; Spence, 1960).

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

GRICE, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination. *J. exp. Psychol.*, 1948, **38**, 1–16.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

KELLER, F. S., & SCHOENFELD, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.

KENDLER, H. H. Learning. *Annu. Rev. Psychol.*, 1959, **10**, 43–88.

MCGUIGAN, F. J., & CROCKETT, F. Evidence that the secondary reinforcing stimulus must be discriminated. *J. exp. Psychol.*, 1958, **55**, 184–187.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

(Received August 16, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 57–64

AMPLITUDE OF RESPONSE AND THE DELAY-RETENTION EFFECT[1]

YVONNE BRACKBILL

University of Colorado Medical School

WILLIAM E. BOBLITT, DOUGLAS DAVLIN, AND JOHN E. WAGNER

University of Denver

In 2 experiments, 3rd-grade boys learned a series of discriminations under reinforcement delay intervals of 0 or 10 sec. and under a condition of low or high amplitude of response. Relearning took place 1 day or 8 days following acquisition. Retention was facilitated by delay for all groups in both experiments. It was also facilitated by high amplitude of motor response, but not by amplitude of verbal response. There was no interaction between reinforcement delay and response amplitude. Amplitude of response varied inversely with learning efficiency. An interpretation of this unexpected finding was offered.

In a recent review of American studies of conditioning, a Soviet scientist pointed out to his readers one marked discrepancy between American and Soviet research. American research, he noted, overemphasizes gross measures of response, such as frequency, and underemphasizes, to the point of neglect, some of the more refined response measures that Soviet scientists have found useful and revealing in the study of behavior. One such measure is the amplitude of a response, i.e., its extent or intensity.

Soviet research on response amplitude has had a long history, beginning with Pavlov's law of strength which states that, other things equal, the more "energy" reaching the cerebral hemispheres, the stronger the resulting conditioned response. (The theory is formulated in terms of the inferred variable, proprioceptive stimulation, rather than in terms of the observed variable, amplitude of muscular response.) The law has been confirmed in essence by much subsequent research, although it has been modified in detail (Asratian, 1953; Popova, 1960). In broad form, it appears as pertinent to operant conditioning and discrimination learning as to classical conditioning, and as applicable to human as to animal learning.

The present study investigates as an independent variable the effect of response amplitude during learning upon subsequent retention of the learned material. If, as has been found, higher amplitude of response facilitates learning, then it might also be expected to facilitate retention. The second variable under investigation, reinforcement delay, has already been found to improve retention (Brackbill, Bravos, & Starr, 1962; Brackbill, Isaacs, & Smelkinson, 1962; Brackbill & Kappy, 1962). Therefore, it was our further hypothesis that the two variables together should combine

---

[1] This investigation was supported in part by Research Grant M-5994 awarded by the National Institutes of Health to the first author.

The authors are very grateful for the cooperation and facilities provided in Exp. I by Rolland G. Walters, Helen Chancellor, and Dorothy W. Wills, Englewood Public Schools; and in Exp. II by John V. M. Hitch, Jane Busselle, and Arvin Lankenau, Aurora Public Schools.

to facilitate retention to a greater extent than either alone.

This study contains two experiments. The learning task is the same in both. It requires *S* to make two sequential responses on every trial—first a verbal response (articulating the name of the stimulus item thought to be correct) and second, a motor response (pressing the microswitch lever adjacent to that item). In Exp. I, groups of *S*s differ in terms of amplitude of the verbal response required of them during learning; they do not differ in amplitude of motor response. In Exp. II, the groups differ in terms of amplitude of motor response, but not in amplitude of verbal response.

EXPERIMENT I

Electromyographic Calibration of Bigram Material[2]

This procedure was preliminary to the main experiment. The *S*s were 5 boys without speech defects. Their average age was 9 yr., 6 mo. They did not serve as *S*s in the main experiment.

The apparatus used was a Grass electroencephalograph, Model III-D. The placement of active electrodes was as follows. Two electrodes were located under the lower lip, near the corners of the mouth (*m. orbicularis oris*). Another two were placed in the submandibular position, approximately halfway between sternomastoid and midline. Two more electrodes were placed on the maseter. However, the maseter records were not used in the data analysis, since they were consistently and markedly inferior to the records obtained from the underlip and submandibular regions.

The material for this preliminary procedure consisted of all two-letter sequences actually occurring in 17,080 written English words. There are 456 such bigrams, according to the Underwood and Schulz (1960, Appendix D) tabulation. The bigrams were spelled out loud to *S* at the rate of one bigram every 5–10 sec. The presentation rate was varied in order to minimize anticipatory muscle movement. The *S* was instructed to repeat the bigram as soon as *E* finished spelling it. The stimuli were administered in a different order to each *S*. To control for adaptation effects, the 30 bigrams first presented were readministered later in the session.

After all *S*s had been run, the five electromyograms were cut into discrete-response records. These were regrouped and recombined on the basis of bigram identity. Thus there were 456 data-units composed of five records each. These were given to three judges, who independently and blindly rated each of the 456 composite records on a 3-point scale of frequency and amplitude of muscular activity. Those bigrams rated high and those rated low by at least two judges were readministered to Pilot *S* No. 1. The records from this session were rated as before, except that unanimous agreement was required of the judges. The final selection included only those bigrams that were unanimously judged as high or low *and* that had been so judged previously by at least two of the three judges. The final selection thus reflected several aspects of reliability: between *S*s, between judges, between testing sessions, and between judging sessions.

As a footnote to the procedure of quantifying EMG recordings, it should be pointed out that the muscle activity involved in articulating a two-letter combination has three components: the activity required to pronounce the first letter, that required for the second letter, *and* the activity referable to the transition between letters. For example, more effort is required for the transition from front vowel to back vowel, as from [i] to [u], than from back vowel to back vowel, as [o] to [u].

---

[2] The execution of this pilot study was made possible by Louise B. Speck, now with the National Institutes of Health, who generously donated facilities, apparatus, and expert advice.

A detailed account of the pilot procedure, a list of the stimulus items used in Exp. I, and the verbatim instructions for both experiments have been deposited with the American Documentation Institute. Order Document No. 7504 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress. The list of stimulus items used in Exp. II can also be obtained through ADI (see Brackbill & Kappy, 1962, p. 15, Footnote 3).

Method

Experimental design.—In this experiment there were two levels of reinforcement delay during learning (0 sec. and 10 sec.), two retention intervals (1 day vs. 8 days between learning and relearning), and two values of response amplitude (low vs. high levels of muscle activity). Thus there are four groups in the analysis of learning scores, and eight groups in the analysis of retention measures. There were 8 *S*s in each experimental group.

Subjects.—The *S*s were 64 third-grade, white, monolingual boys with a median age of 9 yr., 3 mo. The *S*s were obtained from three elementary schools located in the same middle to lower-middle class suburban area. As determined by the California Mental Maturity Test, the median IQs for all third-grade boys at these schools were 105, 105, and 109. All available *S*s were used in two schools; *S*s in the third school were selected at random. The *S*s were randomly assigned to experimental groups.

Experimental task, material, and apparatus.—The experimental task was a series of 18 two-choice simultaneous discriminations. The material for this task consisted of two sets of 18 pairs of bigrams. The articulation of the bigrams in one set required relatively high amounts of activity in the muscles of the throat and mouth. Articulation of the bigrams in the other set was characterized by relatively low amounts of muscular activity. The means of determining amount of muscular activity or amplitude was outlined above.

The two sets of bigrams were matched as closely as possible on all frequency variables, as follows. Between sets: For the 36 high amplitude bigrams, the mean frequency of occurrence in 17,080 written English words was 230.16; the median was 122.50. For the low amplitude bigrams, the corresponding figures for mean and median were 212.83 and 122.50. Between pairs of bigrams: In the high amplitude set, the 18 pairs of bigrams had a median difference in frequency of occurrence of 10.50 out of 17,080 words. For the low amplitude set, the median difference between pairs was 9.00/17,080. Within bigrams: For the high amplitude set, six letters of the alphabet appeared at least five times, while seven letters did not occur once. In the low amplitude set, seven letters of the alphabet occurred at least five times, while seven others were not once represented.

The bigram pairs were randomly ordered within each set, and every successive presentation of the set followed a different random order. The left-right position of the correct stimulus was similarly randomized. The correct bigram within each pair of bigrams was balanced within and between groups.

From *S*'s view, the apparatus was seen as an upright piece of plywood (42 × 14 in.) with two identical columns of parts. Starting from the top, each column consisted of a red pilot lamp, a stimulus aperture, a marble aperture, a marble receptacle, and a lever. A Dixie cup, into which *S* deposited the marbles he earned, sat on the table in front of the apparatus. A standard array of 15 popular 10¢ toys was displayed in another part of the room. An air cooler fan served to ventilate the experimental room and to mask extraneous noises. A small wall mirror was placed at an angle allowing *E* to observe *S*'s manipulation of the apparatus.

By pressing either lever on the apparatus, *S* activated a timer that had been preset for the appropriate response-reinforcement delay interval (0 or 10 sec.). If the lever under the correct picture had been pressed, then, after 0 or 10 sec., the lamp above it flashed on, a loud buzz sounded, and a marble dropped into the receptacle. If, however, *S* had pressed the lever under the incorrect picture, then after the appropriate delay interval, a click sounded, and the lamp situated above the correct picture flashed on. A second timer was used to control the intertrial interval. (The apparatus is described more fully in Brackbill & Kappy, 1962).

Procedure.—The *S*s were run individually in experimental rooms provided by the schools. Each *S* was seen on three separate occasions. The first two sessions were devoted to acquisition and were in all cases held on consecutive days. In the first session *S* learned a pretraining item, followed by Items 2–8; in the second session he learned the remaining items. Retention was tested for one half of the *S*s on the third consecutive day, i.e., 1 day following acquisition, and for the remaining *S*s 8 days following acquisition.

At the beginning of Session 1, *S* was shown to the experimental room where he and *E* examined each of the available toys. The *E* then asked, "Do you see one that you would like to try for?" (The *S* was discarded if he failed to respond positively to this question on this or any subsequent experimental day.) The incentive chosen was removed from the display and placed within easy visual—but not physical—reach of *S*. One pretraining item was used to familiarize *S* with the operations of the machine and to demonstrate the consequences of a correct and an incorrect response. The pretraining item was used until *S* had responded to it correctly on three consecutive presentations, or until he had made three errors. In the latter case, he was discarded. After this, *E* said, "All right, we'll do the same thing with these letters. If you choose the right letter, you will get a marble. And if you can get enough marbles in your Dixie cup, you can have the ——. I will count the marbles you get and tell you when you have enough." On each trial, *S* spelled aloud the letters he thought correct and then pressed the lever corresponding to those letters.

Except for pretraining, the procedure for Session 2—during which *S* learned the remaining 10 items—was essentially the same as for Session 1. On both days, an item was withdrawn from subsequent presentations of that day's set of items after *S* had made three consecutive correct responses to it. Also, for all *S*s, on both days, the intertrial interval was 20 sec. The stimulus item remained exposed during the entire 20 sec.

Session 3 was devoted to relearning. This was carried out under 0-sec. delay and 15-sec. intertrial interval for all *S*s. Relearning scores obtained under these conditions show no significant difference nor systematic differences in trend when compared to relearning scores obtained under a 10-sec. delay interval and a 20-sec. intertrial interval (Brackbill & Kappy, 1962, Control B). The *S*s relearned all items to the same criterion of three consecutive correct responses.

Results

It was necessary to replace five *S*s for the following reasons. One *S* was absent on the day scheduled for his second session. Two were absent from the third session. A fourth *S*—despite a successful pretraining experience in which he learned Item No. 1 to criterion—began his attack on Items No. 2–8 by responding randomly. At the end of 70 trials, he was still responding randomly, and was dismissed. (Random responding in this situation means that *S* chooses

TABLE 1

Mean Retention Scores Obtained from Exp. I and II

Source of Data	Measure of Retention	Response Amplitude	Length of Retention Interval			
			1 Day		8 Days	
			Delay in Learning		Delay in Learning	
			0 Sec.	10 Sec.	0 Sec.	10 Sec.
Exp. I N = 8/group	Trials to relearn	Low	33.75	17.62	41.00	27.12
		High	29.00	13.38	41.38	16.50
	Savings in trials	Low	1.88	23.12	2.00	21.38
		High	11.62	26.50	2.62	29.00
	Errors in relearning	Low	24.00	14.12	29.00	19.62
		High	23.62	11.25	28.75	12.00
	Savings in errors	Low	−8.00	5.25	−9.12	3.88
		High	−.75	8.25	5.62	11.75
Exp. II N = 10/group	Trials to relearn	Low	4.10	1.80	5.20	3.80
		High	3.40	3.10	6.30	4.70
	Savings in trials	Low	11.70	18.10	10.90	14.00
		High	17.40	25.50	16.00	18.20
	Errors in relearning	Low	3.10	1.70	4.10	3.80
		High	2.90	3.10	5.60	4.30
	Savings in errors	Low	10.90	16.30	8.90	11.30
		High	15.90	18.90	14.20	13.70

TABLE 2

F RATIOS OBTAINED FROM FOUR ANALYSES OF VARIANCE OF RELEARNING AND SAVING SCORE MEASURES: EXP. I

Source	*df*	Trials to Relearn	Savings in Trials[a]	Errors in Relearning	Savings in Errors[a]
Amplitude (A)	1	1.60	1.32	1.33	3.51
Delay (D)	1	21.42***	20.51***	25.08***	20.53***
Retention (R)	1	4.48*	<1	2.87	<1
A × D	1	<1	<1	1.05	<1
A × R	1	<1	<1	<1	<1
D × R	1	<1	<1	<1	<1
A × D × R	1	<1	<1	<1	<1
Error (*MS*)	56	(232.09)	(330.77)	(93.29)	(136.17)

[a] A constant value was added to the scores in this analysis.
* $p < .05$.
*** $p < .001$.

the particular lever he is going to press and positions his hand over it *before* the discrimination item is presented.) The fifth *S*, a clinically hyperactive child, was discarded when he began to dismantle the apparatus.

Acquisition.—Low response amplitude material: For the 0- and 10-sec. groups, the mean numbers of trials to criterion were 39.06 and 44.12; the mean numbers of acquisition errors were 17.94 and 21.19. (The term acquisition errors refers to all errors made after the first presentation of the 18 stimuli.) High response amplitude material: For the 0- and 10-sec. groups, the mean trials to criterion were 44.38 and 45.12, while the mean errors were 23.00 and 22.19. Uniformly nonsignificant results emerged from the analyses of variance of trial scores and of error scores; the largest value of F was 1.47 (effect of amplitude on errors).

Retention.—Mean retention scores for the eight experimental groups are shown in the upper half of Table 1. The results of the analyses of variance of these scores are given in Table 2. The first main effect, response amplitude, falls far short of significance in three analyses, and borders on significance in the fourth. Although differences among groups are nonsignificant, the direction of the differences is consistent for the four measures.

The second main effect, amount of delay during acquisition, is highly significant, no matter how or when retention was measured. The third main effect, length of retention interval, is significant in one of the four analyses; in comparison to *S*s tested 1 day after acquisition, those tested 8 days later required more trials to relearn the material but did not make significantly more mistakes. For the saving score data, there are neither significant differences nor systematic trends with respect to length of retention interval.

EXPERIMENT II

Method

Subjects.—The *S*s were 80 third-grade boys drawn from three elementary schools in a common neighborhood. The *S*s were comparable to those of Exp. I in terms of age, race, and socioeconomic background. According to the California Mental Maturity Test, administered and scored by the third-grade teachers, the median IQ for all third-grade boys in the three cooperating schools was 113.5. According to the Kuhlmann-Anderson test, administered and scored by the *E*s, the median IQ for all third-grade boys actually used as *S*s in this experiment was 102.

TABLE 3

F RATIOS OBTAINED FROM FOUR ANALYSES OF VARIANCE OF RELEARNING AND SAVING SCORE MEASURES: EXP. II

Source	*df*	Trials to Relearn	Savings in Trials[a]	Errors in Relearning	Savings in Errors[a]
Amplitude (A)	1	<1	5.48*	2.00	6.63*
Delay (D)	1	4.39*	4.28*	1.53	4.07*
Retention (R)	1	8.09*	2.02	9.59**	5.47*
A × D	1	<1	<1	<1	<1
A × R	1	<1	<1	<1	<1
D × R	1	<1	<1	<1	1.20
A × D × R	1	<1	<1	1.32	<1
Error (*MS*)	72	(8.93)	(114.45)	(6.39)	(44.13)

[a] A constant value was added to the scores in this analysis.
* $p < .05$.
** $p < .01$.

Selection and assignment to groups was random, with the single exception that the eight experimental groups of 10 *S*s each were matched for weight. It had been previously determined from data made available by the Child Research Council, that for 9–10 yr. old boys weight is positively related to grip strength, a motor skill not far removed from the lever pressing operation used in this study. The mean weight for all *S*s was 65.50 lb. The mean weights of the eight groups were matched within a fraction of 1 lb.

Experimental design and procedure.—The experimental design was the same as that of Exp. I. The learning task was also the same: a series of 18 two-choice discriminations. The material for the task was not the same as that used in Exp. I, but rather came from the original experiment (Brackbill & Kappy, 1962); it consisted of line drawings of familiar objects. The apparatus and procedure duplicated that previously described except for the operations defining response amplitude. In this case, a spring was inserted under each of the two microswitch levers on the face of the apparatus. For the low amplitude groups, the springs that were inserted required 2 lb. pressure to depress the lever sufficiently to activate the circuit. For the high amplitude groups, the springs required 4 lb. pressure. Furthermore, *S*s assigned to the 10-sec. delay groups had to hold the lever in its depressed position during the delay interval in order to keep the delay circuit from resetting. The *S*s were allowed to depress the levers with two fingers of either hand, but not with both hands simultaneously nor with fist, elbow, etc. The *S*s sat in a standard size folding chair at a table 29 in. high. The front edge of the apparatus was 11 in. from the front edge of the table.

Results

Seven *S*s had to be replaced for the following reasons. Two *S*s left the state after the first session and before the third. Two others were absent on the day scheduled for their third session. A fifth *S* was discarded because of equipment failure, and the last two *S*s, because they continued for 70 trials to respond randomly.

Acquisition.—For the 0- and 10-sec. groups run under a low amplitude condition, the mean numbers of trials to criterion were 15.90 and 18.75, while the mean numbers of errors were 3.45 and 6.55. For the 0- and 10-sec. high amplitude groups, the mean numbers of trials were 21.55 and 25.60; the mean numbers of errors were 8.80 and 10.05. Response amplitude emerges as a significant main effect in both the analysis of variance of trials, $F\ (1, 76) = 5.05$, $p < .05$, and in the analysis of errors ($F = 8.51$, $p < .01$). In both cases response amplitude is inversely related to learning efficiency.

Both analyses yielded small and nonsignificant *F* ratios for delay and for the interaction of delay with amplitude.

Retention.—The lower half of Table 2 contains the mean retention scores for the eight experimental groups. The results of the analyses of variance of these scores are presented in Table 3. Response amplitude is significant in the two analyses of savings scores, delay is significant in all but the analysis of relearning errors, and length of retention interval is significant in all analyses except savings in trials. The interaction effects are inconsiderable and uniformly so.

General Results and Discussion

Although reinforcement delay affected retention to a significant extent in both experiments, the effect appeared to be less marked in the second experiment than in the first. Any diminution of the effect in Exp. II may well have been due to the distraction, and possibly fatigue, occasioned by the bar holding procedure. In point of fact, it was not an easy task for a 65-lb. boy to exert 4 lb. of pressure with two fingers for 525 sec. per session. Furthermore, *S* was constantly required to attend to or monitor the holding operation, since the slightest decrease of pressure broke the delay circuit, thus requiring a 10-sec. period of steady pressure to be started all over. (This actually occurred only once in 17 trials on the average.)

The second variable under investigation, response amplitude, appears to be independent of the delay retention effect and therefore of little help in explaining the reason for its occurrence. However, the data relating amplitude to retention and to acquisition are of interest in themselves. First, in comparing across experiments, it appears that retention was affected by amplitude of motor response but not by amplitude of verbal response. This finding is similar to that of an early study by Leavitt and Schlosberg (1944), who found that a motor skill was better retained than a verbal skill. The difference, in terms of percent saving scores, was significant for a retention interval as long as 28 days. Of course, the difficulty in making any such comparison lies in the arbitrary basis for equating initial comparability and final extent of mastery of the two types of skill.

Considering now only amplitude of motor response, the results of Exp. II revealed a significant difference between responding at 4 lb. pressure and responding at 2 lb. In view of this, it is pertinent to ask whether there would also be a difference between 2 lb. and 0 lb. This latter condition was not included in Exp. II, but the necessary data, based on an *N* of 24, are available from the original experiment (Brackbill & Kappy, 1962). In order that Exp. II might legitimately be compared to the original experiment, the investigators used in both the same experimental design, task, and procedure. The *S*s in Exp. II, though drawn from a different school system than the original *S*s, were indistinguishable from them in terms of socioeconomic status, age, monolingual home environment, race, and IQ.

From a two-way analysis of variance—with experiments or amplitude as one main effect and acquisition delay as the other—no difference was found between the 2- and 0-lb. conditions on either trials to criterion or acquisition errors. The F ratio for amplitude was less than 1.00 in both cases. The results from analyses of saving scores are quite different. For the effect of amplitude on savings in trials, $F(1, 56) = 5.61, p < .05$. For the effect of amplitude on savings in errors, $F = 53.58, p < .001$. In both cases, the higher response amplitude produced greater savings. Like the effect of delay, the effect of higher amplitude of motor response is to facilitate retention.

Although retention has been the focus of these experiments, some attention must be given to the data relating amplitude to acquisition. In Exp. I, amplitude was found to be unrelated to acquisition measures, and in Exp. II, to be inversely related to learning efficiency. These data are somewhat perplexing in their failure to confirm previous findings indicating that higher amplitudes of response yield greater learning efficiency. Koltsova has found both classically and instrumentally conditioned discriminations to be dependent upon certain parameters of motor response (reported in Brackbill, 1962). For example, in the experiment of Koltsova's most relevant to the present study, infants between 3 and 36 mo. of age were exposed to auditory stimuli in a simple instrumental procedure. Upon presentation of CS+, *S* was to reach for a hidden piece of candy. On CS− trials, no candy was present and *S* was not to respond. Three groups of infants differed in the length of reach or amount of muscular effort, required to get the candy. According to Koltsova, the number of trials taken to establish the discrimination was inversely related to the amount of motor involvement required of *S*.

The reason for the difference between our results relating amplitude to acquisition and those previously obtained may well be attributable to an important methodological difference. In the two experiments of the present study, in contrast to those reported by Koltsova, the experimental task was, relative to the age and ability levels of the *S*s involved, complex and difficult; its mastery requiring *S*'s full attention and mental effort. In addition, however, the act of responding also required attention, since the responses to which "amplitude" referred were not already well routinized aspects of behavior, as was the response of reaching in Koltsova's procedure.[3] In brief, we would propose that amplitude of response may be positively related to efficiency of learning, except for responses which themselves demand attention enough to interfere with learning.

---

[3] Regarding this point it has already been noted that in Exp. II the delay circuit would reset if the pressure on the depressed lever were not sustained and steady. As concerns Exp. I, it should be pointed out that for 9-yr.-old children of only average ability, reading two-letter sequences is a still incompletely mastered skill. Indeed, from pilot work it was found that using trigrams instead of bigrams made the learning task impossibly difficult.

REFERENCES

ASRATIAN, E. A. *Fiziologiia tsentralnoi nervnoi sistemy.* [Physiology of the central nervous system.] Moscow: Medgiz, 1953.

BRACKBILL, Y. Research and clinical work with children. In R. Bauer (Ed.), *Some views on Soviet psychology.* Washington, D. C.: American Psychological Association, 1962.

BRACKBILL, Y., BRAVOS, A., & STARR, R. H. Delay-improved retention of a difficult task. *J. comp. physiol. Psychol.*, 1962, **55**, 947–952.

BRACKBILL, Y., ISAACS, R. B., & SMELKINSON, N. Delay of reinforcement and the retention of unfamiliar, meaningless material. *Psychol. Rep.*, 1962, **11**, 553–554.

BRACKBILL, Y., & KAPPY, M. S. Delay of reinforcement and retention. *J. comp. physiol. Psychol.*, 1962, **55**, 14–18.

LEAVITT, H. J., & SCHLOSBERG, H. The retention of verbal and motor skills. *J. exp. Psychol.*, 1944, **34**, 404–417.

POPOVA, N. S. On certain features connected with the elaboration of conditioned reflexes to sound and light stimuli in dogs. *Pavlov J. higher nerv. Activ.*, 1960, **10**, 84–92.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received August 17, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 65–71

PRACTICE EFFECTS IN FREE AND ORDERED RECALL[1]

KENT M. DALLETT

University of California, Los Angeles

There have been several failures to obtain learning-to-learn (LTL) in immediate recall experiments, while powerful LTL effects are commonly observed in other learning tasks. 2 experiments compared LTL in immediate Free and Ordered Recall as a function of number of presentations (1, 3, or 5) before a single test of recall (Experiment I), or on Trials 1, 3, and 5 when Ss were tested for recall after each of 5 presentations. With a single recall test, only Ordered Recall Ss showed LTL, while with repeated trials both Free and Ordered Recall Ss improved. In each experiment there is a shift in the direction of relatively greater recency effects with practice, but this in itself does not seem to be sufficient to explain all of the LTL observed.

It is typical of several learning tasks used in the laboratory that *S*s improve markedly with practice, even when successive sets of learning materials have no elements in common. Such improvement has been called "learning-to-learn" (LTL) and has been demonstrated in serial anticipation learning (e.g., Ward, 1937), paired-associate learning (Thune, 1951), and in a paired-associate task involving a motor response (Duncan, 1960).

The three tasks mentioned involve repeated trials with the same materials: in experiments in which only a few trials are given with each set of materials, LTL may or may not occur. Murdock (1960) found no LTL when *S*s were given four trials of free recall with five successive 30-word lists, although Meyer and Miles (1953) found considerable LTL in five-trial free recall of nonsense syllable lists. In experiments involving only one trial per list, LTL appears to be still more difficult to obtain. Deese (1957) found no LTL in immediate free recall of word lists. Conrad (1960) found no LTL with digit lists, and Shepard and Teghtsoonian (1961) found no LTL in a similar short-term memory situation in which items are presented intermixed with single-item recognition tests. In these experiments, LTL was not a primary concern of *E*; in other studies practice effects have been sought and found using digit lists presented once (Gates & Taylor, 1925; Martin & Fernberger, 1929; Pollack, Johnson, & Knaff, 1959) and a one-trial paired-associate task (Bills, 1927). Altogether, the contradictory findings in the immediate recall situation contrast strongly with the powerful and obvious effects of practice in standard paired-associate or serial anticipation learning experiments.

Comparison of the experiments mentioned above suggests that one requirement for sizeable LTL effects is that multiple trials be given on each successive list. Duncan (1960), with 10 trials per list found relatively little LTL when only the first trial was examined. Most of the improvement in Duncan's experiment appeared on early trials after the first. The same conclusion is suggested in the data of Meyer and Miles (1953).

---

[1] This research was supported by a grant from the University of California, Los Angeles Faculty Research Committee. The author wishes to express his appreciation for the assistance of Lester D'Andrea who ran most of the *S*s and carried out the statistical analysis of Exp. II. Ola Johnson assisted in the tabulation of the data.

In each of these cases, however, there were clear signs of improvement on the first trial.

Two experiments were carried out in an attempt to explore the effects upon LTL of free vs. ordered recall, multiple presentations (Exp. I), and multiple trials (Exp. II). In the first experiment, *S*s were required to recall 15 successive 10-item word lists by free or ordered recall, and were given one, three, or five presentations of each list before their single test of recall. Since the data suggested that the absence of LTL might depend on the use of a single *test*, even with five presentations, Exp. II was carried out, in which all *S*s were given five *trials* on each of 10 lists, a trial consisting of one presentation plus a test of recall.

METHOD

Materials.—The learning materials were two-syllable words selected from the *Teacher's Word Book* (Thorndike & Lorge, 1944) by taking all words which were (*a*) two syllables in length, (*b*) between 10 and 30 in the general frequency count, and (*c*) different in at least one of the initial two letters from any other word selected. These words were used to construct 15 10-item lists equated for mean and variance of item frequencies. The lists were further equated in intra- and inter-list similarity by eliminating obviously synonymous words (SICKNESS-ILLNESS), and by avoiding within-list duplications of initial letters.

The 15 lists were recorded on magnetic tape, using an Ampex 601 tape recorder. Each list was read five times at a rate of one word per second, with 2 sec. between successive readings of the same list. This tape was one of the two tapes used in both experiments. The second tape was made by duplicating the first tape and splicing it so that the lists (numbered according to their sequence on the first tape) appeared in the order 11, 10, 9 . . . 2, 1, 15, 14, 13, 12.

Experiment I

Procedure.—Each *S* was given three recall sheets, each with space for the recall of five lists. The *S*s were also provided with a cardboard shield which could be used to cover all spaces but the one being used for a given list, and were instructed in its use. Then, *S*s were told that they would hear a series of word lists presented via loudspeaker, and that following each list they would be told, "Now, write," at which time they were to write down as many of the words as possible. Free Recall *S*s were told that they were simply to recall as many words as possible, regardless of order, while Ordered Recall *S*s were told to write down the words in order, drawing a line to indicate the position of any words they could not remember. This feature of the instructions was intended to encourage *S*s to include all the items they could recall, even if one or two in a sequence were missing. A practice list consisting of six three-letter words was read by *E*. Following the practice list, *S*s were played the 15 lists of the experiment with 30 sec. allowed for recall of each.

Half of the *S*s were instructed for Free and half for Ordered Recall. Each of these groups was further divided into three subgroups for whom the lists were presented one, three, or five times before the test of recall. Two tapes were used, making a 2 × 2 × 3 factorial design with 10 *S*s per cell.

Subjects.—The *S*s were 126 students from introductory psychology classes at the University of California, Los Angeles. They participated as part of a class requirement, and were tested in groups ranging in size from one to six persons, with groups being assigned to conditions in rotation. At the end of the experiment, it was necessary to drop some *S*s to balance subgroup *N*s: this was done with the aid of a table of random numbers. A total of 120 *S*s remained.

Experiment II

Experiment II was carried out using the same tapes used in Exp. I. The main difference between the two experiments was that in Exp. II, each *S* was given five trials on each of 10 lists, the first 10 lists on each tape being used. A trial consisted of one presentation of the list followed by a 30-sec. recall test. As in Exp. I, half of the *S*s were instructed for Free, half for Ordered Recall. A total of 43 *S*s was tested in groups ranging from 1 to 3 *S*s in size. After equating groups, there was a total of 40 *S*s remaining in the experiment, 20 in the Free Recall group and 20 in the Ordered Recall group.

RESULTS AND DISCUSSION

For each recall test of each *S*, the number of items correctly recalled was determined without respect to order in recall. In order to keep the analysis of Exp. II parallel to the analysis of Exp. I, only Trials 1, 3, and 5 of Exp. II were analyzed. Since recall of an aurally presented series of words is at least formally similar to articulation testing, there was a small number of items which were judged to be correctly recalled despite the fact that they were not identical with an item which was presented. Examples of such items which were scored as correct were SEATER for CEDAR, LADDER for LATTER, and WEEPER for REAPER. Any difficulties introduced by this ambiguity of scoring are minor, and are uniform throughout each individual experiment. The scoring may have been more lenient in Exp. II in which a phonetically similar item would be considered correct if it were given consistently for the last four or for all five trials.

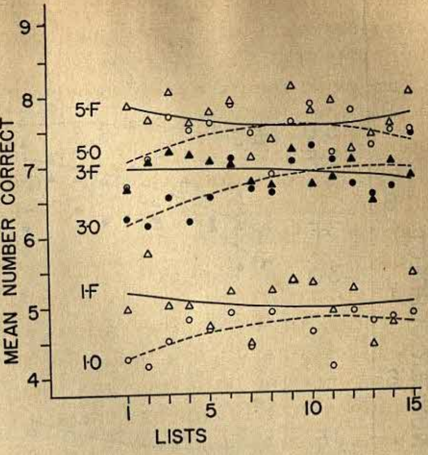


FIG. 1. Mean number recalled, Exp. I.

For each experiment, second-degree polynomial functions were fitted as a measure of LTL. These functions, with the original data points, appear in Fig. 1 (Exp. I) and Fig. 2 (Exp. II). Separate analyses of variance of the linear and quadratic coefficients of Exp. I yielded only two *F* ratios significant at the .05 level. Free and Ordered Recall differ in both the linear, $F(1, 108) = 10.13$, $p < .005$, and quadratic, $F(1, 108) = 8.054$, $p < .01$, components. Individual *t* tests of the linear and quadratic scores for the Free and Ordered groups, pooled over number of presentations, indicate that the trends for Free Recall do not differ significantly from zero—linear: $t(59) = 1.324$, $p > .05$; quadratic: $t(59) = 0.829$, $p > .20$. The corresponding trends for Ordered Recall are highly significant—linear: $t(59) = 3.088$, $p < .005$; quadratic: $t(59) = 3.533$, $p < .005$. By referring to the fitted polynomials in Fig. 1, it can be seen that LTL occurs in Ordered Recall, while the Free Recall groups show no improvement. This pattern appears for each number of presentations, suggesting that multiple trials rather than multiple presentations may be needed to produce LTL in Free Recall *S*s. The analysis of Exp. II tended to support this supposition: in Exp. II there were no differences between Free and Ordered Recall in either linear or quadratic components, and no differences in trend as a function of number of trials. The overall linear trend, however, was highly significant, $F(1, 36) = 15.62$, $p < .001$, suggesting that all *S*s were showing improvement. The two tapes differed in quadratic trend, $F(1, 36) = 12.22$, $p < .005$, in Exp. II, apparently as a result of the relative superiority of performance on Lists 3, 4, 5, and 6 of Tape 2. No such difference appeared in the analysis of Exp. I. The results of both experiments suggest that LTL can occur if *S*s are tested more than once, or if ordered recall is required. The failure to find a summation of these effects in Exp. II might indicate that to some

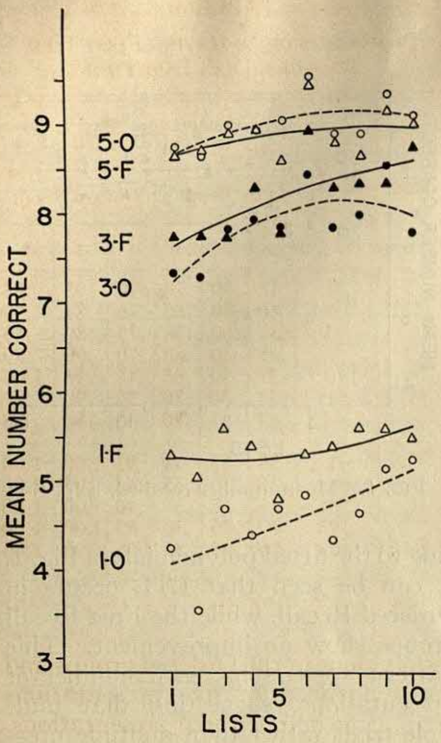


FIG. 2. Mean number recalled, Exp. II.

extent each of these variables produces the same sort of improvement.

Examination of the data points of Exp. II (Fig. 2) shows that while Free Recall *S*s are superior in their performance on Trial 1, Ordered Recall *S*s have surpassed them by Trial 5. A similar finding was reported by Waugh (1961) in a comparison of free and ordered recall involving six trials. While this difference was not statistically analyzed, it will be noted that the crossover appears for 8 of the 10 lists of Exp. II.

Serial position effects and the locus of improvement.—The counterbalancing of lists on the two tapes was such that in each experiment Lists 2 and 3 were identical with Lists 9 and 10. Items correctly recalled were tabulated for these four lists as a function of serial position in presentation. Examining the combined Lists 2 and 3 in each experiment, Ordered Recall *S*s were slightly better than Free Recall *S*s on the first few items of the list, and considerably worse on items near the end of the list. A similar difference was found by Deese (1957), who showed that it results from the relative strengths of items in different serial positions and also from the order in which *S* attempts to recall these items. Free Recall *S*s presumably recall the strongest items first, while Ordered Recall *S*s attempt to recall items in order, beginning with the initial items. Serial position data were obtained in the same way for Lists 9 and 10. In Fig. 3 is plotted the *difference* between the serial position curves for Lists 9 and 10, and those for Lists 2 and 3. In Fig. 3, the results for Trials 1, 3, and 5 are pooled,

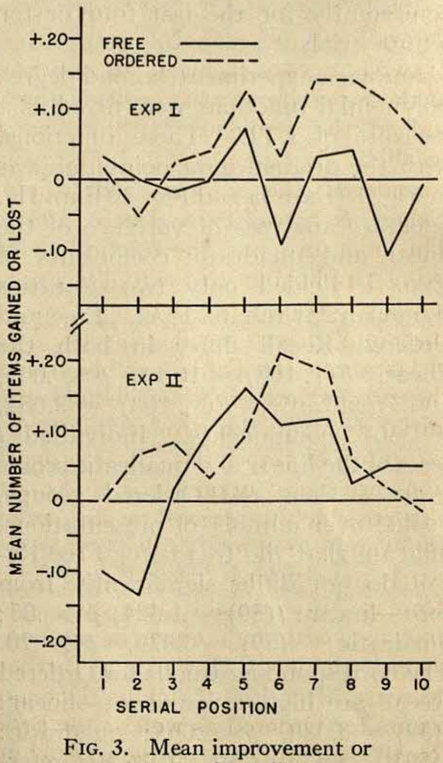


FIG. 3. Mean improvement or loss per *S* per list.

so that each point is based upon 120 opportunities for recall. It will be seen that in Exp. I, Free Recall *S*s show no consistent pattern of change, while Ordered Recall *S*s show improvement toward the end of the list. In Exp. II, there appears to be little improvement at the end of the list: both groups improve at recalling items at or slightly past the middle of the list. Free Recall *S*s in Exp. II seem to have become slightly *worse* at recalling the initial two items.

TABLE 1

PROPORTION OF *S*S HAVING FIRST ITEM STRONGER THAN LAST ITEM

Exp.	Lists	Presentations (Exp. I) or Trials (Exp. II)					
		Free Recall			Ordered Recall		
		1	3	5	1	3	5
I	1	.48	.90	.78	.90	.95	.95
	2	.40	.50	.80	.76	.84	.90
	3	.31	.60	.58	.79	.80	.85
	9	.37	.50	.80	.70	.90	.89
	10	.37	.50	.85	.45	.95	1.00
	11	.31	.50	.70	.76	.90	.89
	13–15	.33	.57	.80	.65	.86	.88
II	1	.55	.55	.44	.95	1.00	1.00
	2	.25	.50	.55	.63	.95	1.00
	3	.21	.37	.35	.90	1.00	1.00
	9	.20	.26	.30	.89	1.00	1.00
	10	.16	.25	.25	.65	.90	1.00

Note.—*N* = 16–20 per cell.

On the basis of these serial position data, it was suspected that for all *S*s the relative strength of initial items in the list was declining relative to the strength of terminal items. This could be described as a shift in the direction of a lesser primacy and a greater recency effect. As a means of evaluating this possibility, the relative strengths of the initial and terminal items in the presentation order were determined for each *S*, using either of two indices of strength. If only one of these two items was recalled, it was judged the stronger. If both were recalled, the item recalled first was considered the stronger (Bousfield, Cohen, & Silva, 1956; Deese, 1957). This measure has the advantage of being applicable to virtually every recall protocol, since it is only rarely that an *S* fails to recall either the first or last item. In Table 1 will be found the proportion of *S*s for whom the initial item was the stronger: with the exception of three cells, these proportions are based upon 19 or 20 *S*s out of a possible 20. It will be seen from an examination of Table 1 that in nearly every instance, there is a decrease in primacy following the first list, although there is no consistent tendency toward a further decrease after List 2. This decrease occurs for Ordered as well as for Free Recall, but appears to be somewhat less evident in the Ordered groups and less evident after five presentations than after one or three presentations. Examination of the serial position data in Fig. 3 indicates that there was little change in whether or not the first item was included in recall except in the case of Free Recall in Exp. II: with this exception, virtually all of the changes reflected in Table 1 are changes in the position of items in recall.

A possible parallel to this effect is provided in an experiment by Anderson and Barrios (1961). In the Anderson and Barrios experiment, *S*s rated their impressions of people after hearing descriptions of these people consisting of six adjectives. The adjectives were arranged in such a way that the relative effects of primacy (the first three adjectives) or recency (the last three) could be judged from the eventual ratings. The relative effect of primacy decreased from the first to the second 20-trial block of the experiment, in

TABLE 2

TOTAL INTRUSIONS FROM PRIOR LISTS: EXP. I

Lists	Free Recall	Ordered Recall
1–3	1	8
4–6	7	6
7–9	8	12
10–12	15	8
13–15	14	5

agreement with earlier results in impression-formation experiments.

Errors.—All errors were classified as being in one of four categories. The first category included responses which were formally similar to an item presented: these were both grammatical variants (SLAUGHTER-SLAUGHTERED), and phonetically similar responses which were too discrepant to have been scored correct. The other categories included (*a*) meaningfully similar responses, (*b*) intrusions from prior lists, and (*c*) other errors—partial responses and items not on any of the lists. Meaningfully similar errors were quite infrequent (a total of 22 in the first experiment, and 9 in the second, taking all five trials into account) as were Other errors. Formal, Meaningful, and Other errors decrease with presentations or trials of the same list, but do not show consistent trends as a function of successive lists. The intrusions from prior lists were quite infrequent in Exp. II (a total of 17), and all but 6 of these originated in the immediately previous list. In Exp. I, intrusions from prior lists were more common: these are presented in Table 2. Again, the major source of these errors was the immediately preceding list, with more remotely previous lists contributing fewer errors the greater their remoteness. For Exp. I as a whole, 47 errors originated in the immediately prior list, 20 from two lists removed, 4 from three lists removed, 7 from four lists removed, and 6 from still more remote lists. The increase in prior-list errors for Free Recall in Exp. I might be associated with the insignificant decline in performance shown by these *S*s as a function of successive lists.

The data on serial position and primacy changes suggest an explanation of the Ordered Recall results in Exp. I. The *S*s instructed to recall items in order began at a disadvantage, since they were not attempting to recall easy items near the end of the list. During the course of the experiment, they increased their recall of terminal items, and of items which, in general, were not in a simple sequence beginning with the first item. This shift from a focus on the beginning of the list to a focus on the entire list could increase their recall scores, and in fact, could bring them to the level of performance at which the Free Recall *S*s had been throughout the experiment. In Free Recall, the shift toward greater emphasis on the end of the list is apparently not effective in changing the mean number recalled. This shift in *S*'s strategy can probably account for all the LTL in Exp. I.

The results of Exp. II do not lend themselves readily to such an analysis. The shift from primacy toward recency is, if anything, more pronounced, but does not strongly affect the *number* of items correctly recalled (Fig. 3): while Free Recall *S*s improve in the number recalled at the end of the list, they show a compensatory decrease in number recalled at the beginning of the list, and in Ordered Recall, there is little change in either initial or terminal items. For each recall method, the main improvement in Exp. II seems to have come at the middle of the list. The improvement with practice in Exp. II might be explained by a modification of Duncan's (1960) suggestion that LTL involves an increase in stimulus discriminability—in this case, a reduction in interfering tendencies which make the center of the list difficult. Such an explanation, in view of the failure of LTL to appear in Exp. I, would need to include an assumption that (*a*) such reduction of interference required more practice than was provided in Exp. I, or (*b*) that such interference appears primarily when *S*s are given several recall tests on the same materials. Certainly the difference between the results of Exp. I and II suggests that multiple recall tests rather than multiple presentations favor the occurrence of LTL.

REFERENCES

ANDERSON, N. H., & BARRIOS, A. A. Primacy effects in personality impression formation. *J. abnorm. soc. Psychol.*, 1961, **63**, 346–350.

BILLS, A. G. The influence of muscular tension on the efficiency of mental work. *Amer. J. Psychol.*, 1927, **38**, 227–251.

BOUSFIELD, W. A., COHEN, B. H., & SILVA, J. G. The extension of Marbe's law to the recall of stimulus-words. *Amer. J. Psychol.*, 1956, **69**, 429–433.

CONRAD, R. Serial order intrusions in immediate memory. *Brit. J. Psychol.*, 1960, **51**, 45–48.

DEESE, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577–582.

DUNCAN, C. P. Description of learning to learn in human subjects. *Amer. J. Psychol.*, 1960, **73**, 108–114.

GATES, A. I., & TAYLOR, G. A. An experimental study of the nature of improvement resulting from practice in a mental function. *J. educ. Psychol.*, 1925, **16**, 583–592.

MARTIN, P. R., & FERNBERGER, S. W. Improvement in memory span. *Amer. J. Psychol.*, 1929, **41**, 91–94.

MEYER, D. R., & MILES, R. C. Intralist-interlist relations in verbal learning. *J. exp. Psychol.*, 1953, **45**, 109–115.

MURDOCK, B. B., JR. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

SHEPARD, R. N., & TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *J. exp. Psychol.*, 1961, **62**, 302–309.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

THUNE, L. E. Warm-up effect as a function of level of practice in verbal learning. *J. exp. Psychol.*, 1951, **42**, 250–256.

WARD, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**(4, Whole No. 220).

WAUGH, N. D. Free versus serial recall. *J. exp. Psychol.*, 1961, **62**, 496–502.

(Received August 20, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 72–77

DEVELOPMENT OF A VERBAL MEDIATOR[1]

MARGARET JEAN PETERSON AND KEITH C. BLATTNER

Indiana University

To investigate the acquisition of a single mediating association 1 paired associate (A-B) was presented, followed by a 2nd paired associate (B-C). On the test trial, S selected the response that seemed "right" to pair with Stimulus A from 3 response alternatives, including the C response. An extraneous syllable was substituted for B in the control conditions. The number of presentations of the establishing stages (1, 3, or 6) and the meaningfulness of the learning materials were varied. Responses reflecting mediational associations and an availability effect were observed. Mediational responses increased as the meaningfulness of the learning materials increased. The number of presentations of the establishing stages influenced the availability of C but not the mediational responding, per se.

Two studies have reported experimentally established mediation using exclusively verbal materials. In the first (Peters, 1935) little evidence of mediation was actually obtained. In the second (Bugelski & Scharlock, 1952) mediated interference was probably produced in the control condition. Bugelski and Scharlock used the customary mediation design in which the response terms of a first list (A-B) serve as the stimulus terms for a second list (B-C). Then a third list is presented which pairs the stimuli from the first list with the responses learned during the second (A-C) so that a chain of associations, A-B-C, may occur. Mediational facilitation was inferred when learning of the third list proceeded more rapidly than learning of a control list in which C responses were paired randomly with A stimuli. However, any mediational responses that were established would probably interfere with test list learning for the control condition, as well as facilitating learning for the experimental condition. Norcross and Spiker (1958) have demonstrated both mediated facilitation and inhibition in addition to a neutral effect using picture pairs as the learning materials. A neutral condition would be produced by the substitution of an extraneous item for B in one of the establishing stages.

This paper reports several studies of the development of a verbal mediating process without the use of successive lists. This was accomplished by testing for mediation (A-C) following presentation of a single A-B pair and a single B-C pair. In the control condition, another item, F, was substituted for B in one of the first two stages. The first three experiments varied the frequency of the initial two associations (A-B and B-C). Meaningfulness of the learning materials was also varied.

METHOD

Design.—The mediational paradigm consisted of three stages, an A-B pair, a B-C pair, and a test trial in which C and two additional alternatives (D and E) were presented with

---

[1] These studies were supported by a United States Public Health Service Grant, M-5209, from the National Institute of Mental Health, United States Public Health Service; the senior author is principal investigator.

The authors wish to express gratitude to Lloyd R. Peterson for his critical reading of the manuscript.

TABLE 1

SAMPLES OF THE LEARNING MATERIALS FOR THE FOUR EXPERIMENTS

Exp.	Stage	Type of Presentation		
		Mediation	Control I	Control II
Exp. I	A-B B-C A-CDE	QJH XZF XZF MWB QJH MWB FHJ GXM	QJH ZGW XZF MWB QJH MWB FHJ GXM	QJH XZF ZGW MWB QJH MWB FHJ GXM
Exp. II	A-B B-C A-CDE	GIL KYT KYT FUS GIL FAM HYP FUS	GIL NOM KYT FUS GIL FAM HYP FUS	GIL KYT NOM FUS GIL FAM HYP FUS
Exp. III	A-B (B-C)[a] A-CDE	GIL heavy (heavy-light)[a] GIL fast soap light	GIL poor (poor-?)[a] GIL fast soap light	
Exp. IV				
High-freq. responses	A-B (B-C)[a] A-CDE	GIL heavy (heavy-light)[a] GIL fast soap light	GIL poor (poor-?)[a] GIL fast soap light	
Low-freq. responses	A-B (B-C)[a] A-CDE	GIL heavy (heavy-lead)[a] GIL fast soap lead	GIL poor (poor-?)[a] GIL fast soap lead	

[a] Assumed to have been established pre-experimentally.

A (A-CDE) as shown in Table 1. To equate frequency of experience with each of the test alternatives, the D and E components for a particular set were C syllables from other sets. C appeared equally often in the first, second, and third position among the test alternatives. Following construction of 36 three-stage mediational sets, two control conditions for each set were prepared by substituting an extraneous syllable, F, for B in either the first (Control I, Column 4, Table 1) or the second stage (Control II, Column 5, Table 1). (Test trials were identical for the three types of presentation of a set.)

To manipulate frequency of presentation the first pair was shown one, three, or six times, following which the second pair was shown with the same frequency as the first. Each cell of the resulting 3 (type of presentation) × 3 (frequency of presentation) design was represented four times on a drum tape. Each set appeared equally often in each of the experimental conditions over the group of *S*s.

The four experiments differed primarily in the meaningfulness of the learning materials (Table 1). CCC trigrams of 0–17% association value (Witmer, 1935) were used in Exp. I and CVC trigrams, of 100% Glaze (1928) association value, in Exp. II.

The CVCs of Exp. II became the A components for Exp. III and IV, while the B components were stimulus words from the Minnesota revision of the Kent-Rosanoff norms (Russell & Jenkins, 1954). The C components were either the most frequently occurring response (35–83%) or a response occurring with low frequency (1–3%) to each of the B stimulus words. Only high frequency responses were used in Exp. III. Both high and low frequency responses were used in Exp. IV to construct two types of mediational presentation, High and Low.

Assuming that the B-C stage had been established during the pre-experimental history of *S*s, the paradigm investigated in Exp. III and IV was actually B-C, pre-experimentally established, then A-B, and the test (Table 1). The associative linkage during the test trial would be the same as in Exp. I and II even though the temporal sequence of the first two stages differed: A was associated with the mediational aspects of B, which had been associated with C, so that A would tend to elicit C. Because B-C preceded the learning of A-B for Exp. III and IV, the results obtained might be considered a conservative estimate of the mediational linkage established.

In the control conditions for both Exp. III and IV, a word, given in response to other Kent-Rosanoff stimuli but judged to have minimal associative strength with C, was substituted for B during the A-B trials. In Exp. III, with the A-B stage but not B-C presented and only one control group, the design became 2 types of presentation × 3 frequencies. Each cell was represented six times among the 36 sets on a tape. The A-B stage was presented once only during Exp. IV so the two types of Mediational presentations, High and Low, and their controls were each represented nine times on a tape. In Exp. III and IV, as in Exp. I and II, each of the 36 sets appeared in every experimental condition.

Procedure and apparatus.—The materials were displayed at a 4-sec. rate on a memory drum. The *S*s spelled aloud the consonant and nonsense syllables or read aloud the English words. The drum was stopped on the test trials while *S* orally paired the A syllable with each of the response alternatives and then reported his choice, the response that seemed "to fit best with the stimulus syllable or to be right." Questions were answered by telling *S*s that the *E* was interested in determining what appeared "right" to the individual *S*. No additional clarification was given except that *E* said, "That's the way," after *S* had made a choice on the practice trials. Three practice trials were given, representing each frequency of presentation once. All practice trials were control trials composed of material not duplicated in the remainder of the experiment. A 4-sec. interval separated the sets.

Subjects.—A total of 132 Indiana University students participated in partial fulfillment of their introductory psychology course requirements, 36 in each of the first three experiments, and 24 in the fourth.

RESULTS

Experiment I, low association value CCCs.—The proportion of C responses for the Mediation condition was .59; for the first control, was .44; and for the second control, .47. An analysis of variance showed the *F* for Type of Presentation to be significant beyond the .001 level, $F\ (2, 70) = 8.40$. The *t*'s between the mean values averaged over the frequencies of presentation shown in the last columns of Table 2 were 6.76 for the Mediation-Control I and 5.30 for Mediation-Control II comparison, both significant beyond the .01 level with $df = 70$. The two controls did not differ reliably ($t = 1.46$). Mediational facilitation seemed clearly demonstrated.

Both control conditions were found to be reliably higher than the .33 level of chance responding ($t_{\text{Control I}} = 3.07$ and $t_{\text{Control II}} = 4.16$, $p < .01$). The higher than chance responding on the

TABLE 2

PERCENTAGES OF C RESPONSES UNDER MEDIATIONAL AND CONTROL CONDITIONS FOR EXP. I–IV

	First Half									Second Half									*M*		
	Mediation			Control I			Control II			Mediation			Control I			Control II			Mediation	Control I	Control II
	1	3	6	1	3	6	1	3	6	1	3	6	1	3	6	1	3	6			
Exp. I Low-association value CCCs	51	54	65	46	46	51	36	46	51	61	54	64	43	36	43	40	50	54	59	44	47
Exp. II High-association value CVCs	68	85	92	65	70	74	67	75	68	72	83	85	67	72	72	70	79	72	79	72	72
Exp. III High-frequency Kent-Rosanoff	80	81	81	34	36	27				83	82	85	32	34	28				81	33	
Exp. IV High-frequency Kent-Rosanoff	81			24						77			37						81	33	
Low-frequency Kent-Rosanoff	74			25						72			35						73	31	

part of the control conditions suggested the presence of a sensitization factor which has been called in the literature response availability.

The F for the Frequency of Presentation factor was significant at the .05 level, $F(2, 70) = 3.23$. Increases in C responding corresponded to increases in the frequency of presentation of the initial stages for all types of presentation (Table 2). The Type of Presentation × Frequency of Presentation interaction term was nonsignificant ($F = 1.85$) indicating that the three trends were essentially parallel. Thus the increases for the Mediation condition must be attributed to increasing response availability rather than to enhanced mediational strength, per se.

An analysis of the number of C responses selected in the first and the second half appears in Table 2. Because of the relatively small number of cases per cell, the halves of the experiment were not treated as a separate factor in the overall analysis of variance. Instead, the differences were evaluated by sign tests. None of the tests reached statistical significance.

Experiment II, 100% association value CVCs.—Mediational facilitation was again demonstrated by the superiority of the Mediation condition over the Control conditions, Table 2, $F_{\text{Type of Presentation}}(2,70) = 4.69$, $p < .05$. The controls again exceeded chance, the t for Control I being 5.01 and for Control II, 5.80, both significant at the .01 level. The Frequency F was significant ($F = 3.72$, $p < .01$) but the nonsignificant F (1.32) for the interaction attested to independence of the trends.

Differences in the C responding during the first and second halves of the experiment (Table 1) were not statistically reliable.

Experiment III, Kent-Rosanoff words.—The most obvious result of the use of Kent-Rosanoff words was to show facilitation under the Mediation condition while reducing the level of C responding for the Control to the chance level. The difference between the two types of presentation was significant beyond the .001 level, $F(1, 70) = 88.51$. The F for Frequency of Presentation was less than unity and the interaction $F(1.76)$ was nonsignificant, indicating that increasing the frequency of presentation of the A-B stage did not reliably influence the level of C responding for either of the types of presentation.

Differences between the first and second half were not reliable as evaluated by sign tests.

Experiment IV, high and low frequency Kent-Rosanoff words.—The analysis of variance showed that differences between the types of presentation were significant beyond the .001 level, $F(2, 70) = 51.70$, with both Mediation conditions showing facilitation over their controls, $t_{\text{High-Control}}(46) = 9.45$, $t_{\text{Low-Control}}(46) = 7.94$, $p < .01$. Although the difference was in the expected direction, mediational responding with high frequency Kent-Rosanoff response words did not reliably exceed that learned with low frequency Kent-Rosanoff words, $t = 1.51$. Reliable practice or interference effects between the two halves of the experiment were not observed.

DISCUSSION

Generally, as the meaningfulness of the learning materials increased, mediational responding increased. The negligible difference between the 100% association value CVC trigrams and the Kent-Rosanoff words may have resulted from the direction of the associational chain as manipulated in Exp. III and IV.

In those two experiments, the B-C stage was assumed to have been acquired before the A-B stage was experimentally presented. Or the particular sample of students selected may not have been representative of the norm group used in the Minnesota standardization.

The reliable increase of control responding for 100% association value CVCs relative to that for the low association value CCC trigrams indicated that response availability may be a function of the meaningfulness of the materials. Peterson, Peterson, and Miller (1961) demonstrated that retention of three-letter units over short intervals of time was related to the meaningfulness of the units. In the present experiments, the probability of correct retention would be expected to increase as the materials increased in meaningfulness and the correctly remembered items would be more available for selection on the test trial. This prediction would be made for all conditions of both Exp. I and II and depends on exposure to the C syllable during the second stage of the paradigm.

While the basic conception of response availability is the same, the underlying mechanism responsible for one response being more available than another may be different with associations acquired in the natural language community of *S*s. Rather than resulting from exposure to the item as experimentally presented, the C response is assumed to occur as an implicit response mediationally chained to B even though B and C are not actually presented in juxtaposition. For example, in the Russell and Storms (1955) study of implicit verbal chaining in which they used Kent-Rosanoff words, test list acquisition could have been facilitated by increased availability of the responses without a specific associative connection between A and the test response. The same possibility exists in Exp. III and IV, although McGehee and Schulz (1961) found no evidence of heightened response availability in their carefully controlled replication of the Russell and Storms experiment. Assuming superiority of the Mediation condition over the controls, the question concerns the actual links of the associational chain, not its existence.

The obvious conclusion arising from the ambiguities inherent in the use of natural language habits is that the clearest analysis of a mediational process may be obtained by experimentally controlling experience with the materials to be learned. Separate estimates of mediational responses and of response availability may be checked for artificiality with habits acquired outside the laboratory. This was the primary reason for the third and fourth experiments in this paper.

The data indicated that mediational associations were established by a single presentation of the first stages and that the apparent increases associated with additional presentations reflected increased response availability. However, mediational responses increased as the meaningfulness of the learning materials increased and meaningfulness has often been defined in terms of the frequency with which associations are given to a stimulus. Underwood and Schulz (1960) showed that response members of paired associates, particularly, profit from increased, experimentally controlled experience. It is possible that the mechanisms involved in the actual establishment of a mediating process may be distinct, following a pattern of one-trial learning, or that some artifact of the experimental procedure, such as boredom with the repetitious presentation of the same pairs, affected the obtained results. The simplicity of the design used herein should facilitate evaluation of such theoretical implications when appropriate modifications have been made in the procedure.

REFERENCES

Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediation association. *J. exp. Psychol.*, 1952, **44**, 334–338.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

McGhehee, N. E., & Schulz, R. W. Mediation in paired-associate learning. *J. exp. Psychol.*, 1961, **62**, 565–570.

Norcross, K. J., & Spiker, C. C. Effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 129–133.

Peters, H. N. Mediate association. *J. exp. Psychol.*, 1935, **18**, 20–48.

Peterson, L. R., Peterson, M. J., & Miller, A. G. Short-term retention and meaningfulness. *Canad. J. Psychol.*, 1961, **15**, 143–147.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216, Office of Naval Research.

Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

Witmer, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–360.

(Received August 20, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 78–83

ATTITUDES OF JUDGMENT AND THE SIZE-DISTANCE INVARIANCE HYPOTHESIS[1]

WILLIAM EPSTEIN

University of Kansas

Judgments of the size and distance of standards which were 10, 20, 30, 60, or 120 ft. from O were obtained with the phenomenal, objective, perspective, and projective attitudes. Different Os were assigned to each of 20 combinations of standard distance and attitude. The main purposes of the study were: (a) to verify Carlson's (1962) results concerning the effects of attitude on size judgment; and (b) to determine the effect of attitudes on the size-distance relationship when the possibility for successive comparisons by the same O at different distances is eliminated. The main findings were: (a) The phenomenal, objective, perspective, and projective attitudes produced size matches which were veridical, overestimations, greater overestimations, and underestimations, respectively. (b) Deviations of size judgments varied with distance. (c) Distance judgments did not vary for the different attitudes of size judgment. However, accuracy of distance judgment varied with the distance judged. (d) Size and distance judgments were not related systematically. These findings were discussed in the context of the size-distance invariance hypothesis.

A recent review of the experimental literature (Epstein, Park, & Casey, 1961) has shown that there is a considerable body of experimental data which is inconsistent with the size-distance invariance hypothesis. In particular there is the frequent finding that apparent size increases with increasing physical distance while apparent distance is only rarely found to increase more rapidly than physical distance. In a series of investigations Carlson (1960, 1961, 1962) has reported evidence that overestimation in size judgments is due to an attitudinal bias induced in *O* inadvertently by *E*'s instructions. Carlson argues that instructions which emphasize objective physical equality induce *O* to assume a perspective attitude. This attitude of judgment requires that a nearer object must *appear* to be larger than a distant object if the two are equal in physical size. Therefore, the perspective attitude would lead to larger settings of a variable when the variable is nearer and smaller settings when the variable is farther than the standard. These are the two kinds of matches which provide the data we label overestimation of size. While these considerations would account for overestimation they do not account for the usual finding that overestimation increases systematically with distance. Carlson (1960) explains these results by asserting that in a single experimental setting "overestimation may be a fairly precise function of distance, but only because trials at different distances are not really independent, and *O* can judge the distances relative to each other" (p. 201).

Carlson's experiments are important in the context of the difficulties confronting the invariance hypothesis and, therefore, they deserve careful scrutiny. A review of the complete series of studies (Carlson, 1960, 1961,

---

[1] The work reported in this paper was supported by Research Grant MH 04153-03 from the National Institute of Mental Health of the United States Public Health Service.

1962) produced three observations: (*a*) In all of the experiments the standard was at 40 ft. while the variable was at 10 ft. Since only one pair of distances was used we do not know how variations in distance would affect size judgments under the various attitudinal conditions. (*b*) The same *O* was used under all the attitudinal conditions. Thus, in one experiment (Carlson, 1962) each *O* made judgments under four sets of instructions. It is likely that *O*'s response to a given set of instructions is conditioned by his previous experience with other instructions. Carlson's comments on this point do not eliminate this concern. (*c*) Carlson does not provide an adequate analysis of the data for distance judgment. In the most recent study (Carlson, 1962), the data are not reported at all. The theoretical relevance of these data has been suggested in an earlier analysis of Carlson's work (cf. Epstein et al., 1961, p. 479).

The purpose of the present investigation was to examine Carlson's thesis in an experimental design which took into account the observations recorded above. The design also provided conditions for testing Carlson's assertion that overestimation which increases with distance is a consequence of the opportunity for making successive size judgments at different distances in the same setting.

METHOD

General plan.—Each *O* provided judgments of size and distance at one standard distance and with one set of instructions only. Five distances of the standard and four sets of attitude-inducing instructions were used making a total of 20 combinations of experimental conditions.

Observers.—The *O*s were 200 undergraduates who were fulfilling a course requirement. Ten *O*s were assigned to each of the 20 experimental combinations.

Materials.—The standards were two isosceles triangles cut from white cardboard. The altitude of the small triangle was 7.6 cm. and the altitude of the larger was 15.9 cm. When a standard was presented for judgment it was mounted on a black background which was sufficiently large to obscure the immediately surrounding background of the corridor. The variable was a triangle on an identical black background. The base and altitude of the variable could be varied continuously in the direction of increasing or decreasing size.

An additional device was constructed for the purpose of obtaining judgments of the distance of the standard. This was a specially prepared ruler calibrated in centimeters upon which two identical triangular markers were mounted. The position of the near marker was fixed while the position of the second marker could be varied by sliding it along the ruler. The position of the variable marker could be determined directly by reading the calibrated scale. The scale markings were on the underside of the ruler and were not visible to *O*. There were no distinguishing marks on the side which was visible to *O*.

Experimental conditions.—Combinations of five standard distances and four sets of attitude-inducing instructions comprised the 20 experimental conditions. The standards were located 10, 20, 30, 60, or 120 ft. from *O* at approximately eye level. The variable was always 5 ft. from *O*, and it was separated laterally from the standard by an S-*O*-V angle of 20°. The instructions were adopted *verbatim* from Carlson (1962, pp. 69–70) and, therefore, they will not be repeated here. The four sets of instructions were intended to induce the phenomenal, objective, perspective, and projective attitudes, respectively. Under all conditions *O* made his judgments binocularly. A chin rest was provided for support, but lateral head movements were permitted. The corridor was normally illuminated.

Procedure.—There were two stages in the experiment. In the first stage *O* was given the size-matching instructions. Then *O* made four judgments of each of the two standards. The order of ascending and descending trials and the order of presentation of the small and large standard was counterbalanced. No time limits were imposed on *O*'s performance.

In the second stage of the experiment *O* made two judgments each of the distance of the two standards. Again, the order of judgments was counterbalanced. While making these judgments the variable triangle was in view and was set by *E* at the mean of

TABLE 1

Mean Size Matches with Distance of Standard and Attitude of Judgment Variant: Comparison at Constant Distance of 5 Ft. from *O*

Attitude	Size (cm.) of Standard	Distance of Standard (ft.)									
		10		20		30		60		120	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Phenomenal	7.6	6.91	.73	8.02	.66	8.02	1.50	6.67	1.59	6.95	3.13
	15.9	14.65	1.37	15.42	1.04	16.48	2.05	15.13	3.56	15.59	4.86
Objective	7.6	8.56	1.14	8.55	.93	9.04	1.65	9.71	2.02	12.01	1.53
	15.9	17.04	1.79	15.94	1.42	17.32	2.18	19.23	2.58	23.49	3.48
Perspective	7.6	10.50	1.94	11.75	2.82	12.91	3.18	12.39	4.11	13.99	5.23
	15.9	19.29	2.13	19.14	2.54	20.25	2.84	21.80	4.93	24.86	5.80
Projective	7.6	5.17	1.10	4.46	1.49	3.86	1.43	2.78	1.47	2.48	1.31
	15.9	11.14	2.44	9.55	1.93	7.96	2.15	5.71	3.51	5.01	2.53

O's settings for size. The ruler described above was given to *O* and he was instructed to "adjust the location of the second marker, so that its position on the ruler is in the same ratio to the stationary marker as the far triangle in the hall is to the near triangle."

The experiment was concluded with an interview which examined *O*'s understanding and compliance with the instructions.

RESULTS

Judgment of size.—Table 1 contains the mean judgments of size for the 20 experimental conditions. A comparison between the means in each column separately provides a check on Carlson's (1962) results. Comparing the means within each row permits us to determine whether systematic changes in size matches occur with increasing distance in the absence of the opportunity for successive judgments by a single *O* at different distances. In addition, we can determine whether changes in size judgments with increasing distance are a function of attitude. Of particular interest are the results for the phenomenal and objective attitudes.

For purposes of analysis each *O*'s size judgments were expressed as percentage deviations from the objective size of the standard and an analysis of variance was performed on the coded deviation scores for both standards combined. Figure 1 presents a series of curves which show the percentage deviation in size judgments. Table 2 contains a summary of the analysis of variance of these data. The effect of attitude was significant. Table 1 and Fig. 1 show that the phenomenal attitude yielded the most veridical matches. The matches for the objective attitude were overestimations. The over-

TABLE 2

Analyses of Variance of Coded Percentage Deviations of Judgments of Size and Distance from Objective Size and Distance

Source	*df*	*F*	
		Size Judgments	Distance Judgments
Attitude (A)	3	160.59**	0.26
Distance (D)	4	2.13	29.64**
A × D	12	4.57**	1.65**
Error (*MS*)	180	(540.19)	(285.87)

** $p < .01$.

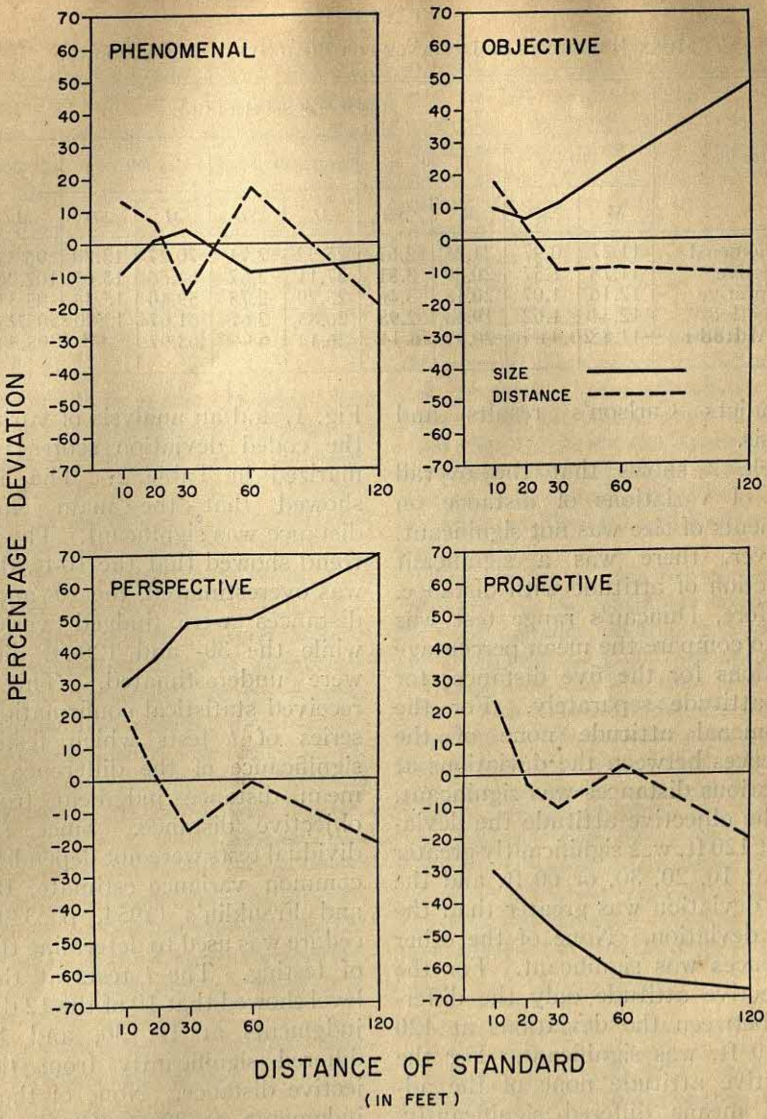


FIG. 1. Mean percentage deviations of size and distance judgments from objective size and distance for both standards combined.

estimation was even greater for the perspective attitude. For the projective attitude considerable underestimation was obtained. Duncan's new multiple-range test (Edwards, 1960, pp. 136–140) was used to compare the mean percentage deviations for the four attitudes at each distance. At 10, 60, and 120 ft. all the differences were significant at the .05 level. At 20 and 30 ft. only the difference between the phenomenal and objective attitude fell short of significance. These results are in substantial agreement with Carlson's (1962) findings although the absence of a consistent difference between the phenomenal and objective attitude

TABLE 3

MEAN JUDGMENTS OF DISTANCE FOR BOTH STANDARDS COMBINED

Attitude	Distance of Standard (ft.)									
	10		20		30		60		120	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Phenomenal	11.27	0.97	21.25	2.67	25.32	2.71	70.27	13.20	96.35	24.01
Objective	11.78	1.57	20.80	2.81	27.11	2.67	64.56	15.65	107.06	26.83
Perspective	12.16	1.07	20.18	3.46	25.29	2.73	59.40	15.38	95.34	22.13
Projective	12.34	1.62	19.43	2.98	26.85	2.67	61.67	13.80	95.25	18.78
All Attitudes	11.88	—	20.41	—	26.14	—	63.97	—	98.50	—

contradicts Carlson's results and analysis.

Table 2 shows that the overall effect of variations of distance on judgments of size was not significant. However, there was a significant interaction of attitude with distance. Therefore, Duncan's range test was used to compare the mean percentage deviations for the five distances for each attitude separately. For the phenomenal attitude none of the differences between the deviations at the various distances was significant. For the objective attitude the deviation at 120 ft. was significantly greater than at 10, 20, 30, or 60 ft. and the 60-ft. deviation was greater than the 20-ft. deviation. None of the other differences was significant. For the perspective attitude only the difference between the deviations at 120 and 10 ft. was significant. For the projective attitude none of the adjacent means differed significantly, however, all of the differences between nonadjacent means, e.g., 120 vs. 30 ft., 30 vs. 10 ft., were significant.

Judgment of distance.—Table 3 contains the mean distance judgments. In analyzing these data the two standards were combined, and each *O*'s average setting was converted into the percentage deviation of his average setting from objective distance. The mean percentage deviations for the 20 groups are plotted in Fig. 1, and an analysis of variance of the coded deviation scores is summarized in Table 2. The analysis showed that the main effect of distance was significant. The overall trend showed that the 10-ft. distance was overestimated, the 20- and 60-ft. distances were judged veridically, while the 30- and 120-ft. distances were underestimated. This trend received statistical confirmation by a series of *t* tests which tested the significance of the difference of the mean distance judgment from the objective distance. Since the individual tests were not dependent on a common variance estimate, Bennett and Franklin's (1954, p. 339) procedure was used to determine the level of testing. The *t* tests at the .005 level showed that 10 of the 12 distance judgments at 10, 30, and 120 ft. differed significantly from the objective distance. None of the mean judgments at 20 or 60 ft. differed significantly from the objective distance.

The judgments of distance made by *O*s in the different attitudinal conditions did not differ significantly. This can be observed by comparing the means within each column of Table 3. Of course there is no reason to expect differences since the attitude-inducing instructions were concerned with size and not distance. However, if it may be assumed that the judgments of distance represented the perceived distance during the time of size judgment as well as on the immediate occasion of distance judgment, then the results are theoretically significant. They indicate that attitudes of observation have powerful effects on size judgments without the mediation of modifications in judged distance. The independence of size and distance judgments is reflected in the curves shown in Fig. 1.

Verbal reports.—The *O*s who were given the objective and phenomenal instructions had no difficulties understanding what was required. Postexperimental questioning revealed that the distinction between phenomenal and physical equality was recognized and utilized in making their judgment. On the other hand, the instructions for the perspective and projective attitudes often had to be supplemented by additional explanation. Five *O*s who were assigned to the projective instructions misunderstood the instructions and made phenomenal judgments. They were replaced and were not included in the data. Most *O*s under the projective instructions complained that while they knew that the variable had to be made smaller, they had no way of knowing how much smaller it should be.

The instructions for distance were well understood by all *O*s.

DISCUSSION

There was unambiguous confirmation of Carlson's finding that the occurrence of overestimation depends on the presence of an attitudinal bias. We observed that the phenomenal attitude led to size matches which approximated constancy. In fact, *t* tests which compared the mean size judgments shown in Table 1 with the actual physical sizes of the standards showed that none of the 10 differences was significant. On the other hand, the objective attitude produced overestimations and the perspective attitude led to still greater overestimation of size.

Despite the support of Carlson's thesis provided by these results, the data were not in full agreement with Carlson's analysis. In particular, there are the findings that under the objective and perspective attitude amount of overestimation increased with increasing physical distance. The percentage overestimation at 120 ft. was approximately five times greater than at 10 ft. for the objective attitude and two and a half times greater for the perspective attitude. In general, there was a consistent trend for percentage overestimations to increase with increasing physical distance (see Fig. 1). Since the opportunity for successive comparison at different distances was eliminated, we are left without an obvious explanation of this relationship.

The results of this experiment have relevance for the continuing evaluation of the size-distance invariance hypothesis. They suggest that the invariant relation between size and distance is a special case which obtains under certain conditions only. The identity of these conditions remains to be specified. However, it does seem likely that these conditions will be characterized by at least two features: (*a*) Judgments will be spontaneous responses unmediated by cognitive deliberative considerations, and (*b*) size and distance responses will be concurrent aspects of the same judgment.

REFERENCES

BENNETT, C. A., & FRANKLIN, N. L. *Statistical analysis in chemistry and the chemical industry.* New York: Wiley, 1954.

CARLSON, V. R. Overestimation in size-constancy judgments. *Amer. J. Psychol.*, 1960, **73**, 199–213.

CARLSON, V. R. Effects of sleep-deprivation and chloropromazine on size-constancy judgments. *Amer. J. Psychol.*, 1961, **74**, 552–560.

CARLSON, V. R. Size-constancy judgments and perceptual compromise. *J. exp. Psychol.*, 1962, **63**, 68–73.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, 1960.

EPSTEIN, W., PARK, J., & CASEY, A. The current status of the size-distance hypothesis. *Psychol. Bull.*, 1961, **58**, 491–514.

(Received August 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 84–90

REVERSAL LEARNING IN RATS AS A FUNCTION OF PERCENTAGE OF REINFORCEMENT AND DEGREE OF LEARNING[1]

ALBERT ERLEBACHER

Northwestern University

An experiment was performed investigating the effect of percentage reinforcement and degree of overlearning on reversal learning in rats. 168 female albino rats, 75–80 days old, were run in a black-white discrimination task according to a 2 × 4 orthogonal design. Acquisition was under 1 of 2 percentages of reinforcement—100% and 50%. Upon reaching criterion the groups were divided into 4 subgroups receiving 0, 70, 140, or 280 trials overlearning. Following overlearning, reversal learning was begun. Acquisition level was also controlled. Results showed that following 50% reinforcement, there is an inverse relationship between the number of overlearning trials and ease of reversal learning. Following 100% reinforcement, no effect due to overlearning was found. Reversal was easier following 100% than following 50% reinforcement. It was concluded from the reversal results that at least 2 factors are required to explain the partial reinforcement effect and the overlearning reversal effect.

The purpose of this research was to investigate the effects of percentage reinforcement and degree of learning upon reversal learning in rats.

At the mammalian level, the weight of the published evidence seems to indicate that there is a direct relationship between degree of overlearning and the ease of learning a new, antagonistic task. In animal reversal learning studies, Reid (1953); Pubols (1956); Capaldi and Stevenson (1957); Bruner, Mandler, O'Dowd, and Wallach (1958); Birch, Ison, and Sperling (1960); Brookshire, Warren, and Ball (1961); and D'Amato and Jagoda (1961) have shown that there is a direct relationship between the amount of overtraining and the ease of reversal learning in rats.

One interpretation of these results is that the increased number of trials in the first task makes the new task more discriminable; i.e., makes the proper mode of responding stand out such that *S* can more easily react to a shift. This view is held by Stevenson and Moushegian (1956) in an explanation of similar results at the human level, and in this paper will be referred to as the *discrimination hypothesis*.

It was the intent of this research to concentrate on the discrimination hypothesis. This hypothesis is highly reminiscent of a hypothesis suggested by several researchers (e.g., Grant & Schipper, 1952; Humphreys, 1939) to account for the increased resistance to extinction following partial reinforcement. Here, the discrimination hypothesis states that the higher the percentage of reinforcement during acquisition, the more likely is *S* able to discriminate the change to extinction.

---

[1] This paper is based upon a dissertation submitted to the Graduate School of the University of Wisconsin in partial fulfillment of the requirements of the PhD degree. The research was performed while the author held a predoctoral fellowship from the National Institute of Mental Health. The author wishes to thank L. E. Ross and E. J. Archer under whose direction this investigation was conducted.

It is obvious that the learning to respond on the reversal task involves the extinction of the previously correct response. Experimental evidence by Birch et al. (1960) shows that the facilitation due to overlearning is accounted for by the differential extinction of first-task responses on the part of the various degree of learning groups. Moreover, it has been shown that the overlearning effect also holds for simple extinction (Murillo & Capaldi, 1961; Senko, Champ, & Capaldi, 1961; Stimmel & North, 1959).

Given the similarity of results in reversal learning experiments and experiments on extinction; and assuming that the discrimination hypothesis can be used to account for both the overlearning effect in reversal situations (which involve extinction) and for the partial reinforcement effect in extinction, one is led to the following hypothesis: When a reversal learning experiment is performed while varying degree of learning and percentage reinforcement simultaneously, each variable should have the same effect (in direction but not necessarily in magnitude) upon performance under all levels of the other variable. That is, the overlearning effect should occur under both 100% and 50% reinforcement and the partial reinforcement effect should occur under all overlearning conditions. This hypothesis has been tested and not confirmed recently on the human level (Erlebacher & Archer, 1961). Perseveration (responding in the *formerly* correct way) decreased with an increase in the degree of first-stage learning only when this learning was under continuous reinforcement, while under partial reinforcement there was a direct relationship between amount of perseveration and degree of first-stage learning.

Erlebacher and Archer (1961) worked with a card sorting task. In the present experiment, an attempt was made to determine if their results obtain at a different level of task complexity. A simple analog of the card sorting experiment is the reversal learning situation. As such an analog, reversal learning should be appropriate for testing the hypothesis.

METHOD[2]

Subjects.—The *S*s were 168 experimentally naive female albino rats purchased from Sprague Dawley. They were 75–80 days old upon arrival in the laboratory.

Apparatus.—The experiment was carried on in an air conditioned room in which the air conditioner fan served as a slight masking noise. The *S*s were carried to the room and held between trials in individual compartments of a carrying box. Pretraining took place in individual compartments of a flat gray box.

The experiment proper was performed with a single unit wooden T maze. The maze consisted of a 10¾ × 2¼ in. start box, a 9¾ × 3½ in. runway, and arms that measured 27 × 3½ in. The entire maze was 4 in. high, and the top was made of glass. The various compartments were separated by guillotine doors which were closed behind *S* as it entered the next section. The arms had a guillotine door 12 in. from the terminal end, constituting the goal box. Each side of the T had two interchangeable arms—one painted flat white, the other flat black. On any one trial a black arm was slid into position on one side while a white one was slid into position on the other. The start box and runway were painted flat gray. The maze was illuminated by four 150-w. Sylvania Flood bulbs, supplied by 25 v., which were placed symmetrically 37 in. above the maze top.

Design.—Three replications of a 2 × 4 × 3 orthogonal design were carried out. The three variables and their values were: percentage of reinforcement (50% or 100%); degree of overlearning (0, 70, 140, or 280 trials); and level of acquisition (fast, intermediate, or slow). In each replication 2 *S*s were run in each fast and slow acquisition cell and 3 *S*s were run in the intermediate acquisition cells. Thus, 56 *S*s were run in each replication, making a total of 168 *S*s for all three replications. The total number of *S*s run for each percentage of reinforcement-degree of overlearning condition was therefore 21.

---

[2] A more detailed presentation of method is given in Erlebacher (1961).

The "levels" variable was introduced in order to control for any correlation that might exist between speed of acquisition and speed of reversal learning. In each replication, *S*s were ordered according to their acquisition speed. The first 8 were randomly assigned to the four overlearning trial groups, 2 per group; the next 12 were then randomly assigned to the four groups; and the last 8 were similarly assigned. This randomization was accomplished separately within each percentage of reinforcement condition since *S*s under, e.g., 50% reinforcement could not have been assigned to a 100% group. To qualify as an orthogonal design, therefore, the assumption was made that the fastest *S*s under 50% would have been the fastest *S*s if run under 100%, etc.

One *S* did not complete acquisition after 350 trials in the second replication. This *S* was discarded, an extra *S* was run in the third replication, and the slowest *S* in the third replication replaced this discarded animal.

Procedure

Pretraining.—On the day of their arrival (Day 1), *S*s were placed on a deprivation schedule. Each *S* was given 45 min. access to Purina lab chow at the same time each day in separate feeding cages. Water was constantly available in both feeding and living cages. Handling *S*s daily while placing them into the feeding cage and replacing them into living cages also served the purpose of taming.

On Days 6–8 pretraining took place and on Day 9 the experiment proper began. Pretraining and all other experimental sessions were always begun 45 min. prior to the feeding time.

Pretraining sessions were carried out in order to train animals to accept the pellets given as rewards in the experiment. A total of 13 pellet eating trials were given on Days 6–8 in the pretraining compartments.

Acquisition.—During acquisition, the white side was always correct for each animal. The *S*s were run 10 trials per day except on the first 2 days of acquisition when they were run only 5 trials per day.

On any trial *S* was taken from the carrying cage and placed in the start box. When *S* faced the runway door, this door was opened and *S* was allowed to run to each successive compartment of the maze, with no retracing permitted. If *S* chose the correct side, two pellets were in the food cup on reinforced trials. On all trials *S* was allowed to remain in the goal box for a period of 5–10 sec., or until the food was ingested. The intertrial interval was set at a minimum length of 3 min.

The *S*s in the 100% reinforcement group were rewarded on every trial on which they chose the correct side. The *S*s in the 50% reinforcement group were rewarded for 50% of their correct choices. A reinforcement schedule was arranged for the 50% group such that on each block of 10 correct choices half of the choices were reinforced and half went unrewarded. Within the 10-choice block, the schedule was randomized with the restriction that there be no more than 3 rewarded or unrewarded correct choices in a row. The entire reinforcement schedule consisted of 10 such successive blocks. One further restriction on the schedule was that the very first correct response in the first block be reinforced. After *S* had made 100 correct responses, the same reinforcement schedule was started again.

The position of the white arm was varied from trial to trial according to the following series taken arbitrarily from Gellerman (1933): LRRLLRLLRR and RLLRRLRRLL. The two series were used alternately on successive days. The *S*s were run to a criterion of 10 correct responses in a row, which could span two sessions. The last criterion trial was always reinforced even when the schedule for the 50% *S*s did not call for a reinforcement. The last criterion trial served as the last trial for that day.

Overlearning and reversal.—After having reached the acquisition criterion, *S*s were assigned to one of the four overlearning conditions. The procedure for all overlearning groups was exactly the same as during acquisition. The 50% group continued to be reinforced for only half of the correct choices, except that the last overlearning choice, if correct was reinforced for all *S*s.

Reversal learning began on the day following the last overlearning session. The procedure during reversal learning was exactly the same as during acquisition except that white was no longer reinforced but black was. Reinforcement during reversal learning was 100% for all *S*s. The *S*s in the 0-trial overlearning groups began reversal learning on the day following the last acquisition trial. Reversal learning was continued until an *S* had reached a criterion of 10 consecutive correct responses.

RESULTS

Acquisition

The data for acquisition were analyzed in terms of total trials to criterion. The means for 100% and 50% reinforcement were 27.8 and 33.8, respectively, $F\ (1, 162) = 17.49$, $p < .01$, showing that acquisition was more difficult under partial reinforcement. The three acquisition levels were, of course, chosen *post hoc* and therefore do, by necessity, differ. The three levels had means 20.1, 29.5, and 43.4, $F\ (2, 162) = 77.04$, $p < .01$. The mean square for error was equal to 85.41.

Reversal

Trials to criterion.—The reversal data in total trials to criterion are presented in Fig. 1 for the various overlearning conditions, with percentage of reinforcement as the parameter. Figure 1 shows that following 100% reinforcement there was no consistent effect due to the number of overlearning trials. After 50% reinforcement, however, reversal became increasingly more difficult as the degree of overlearning increased. Table 1 presents a summary of the analysis of variance for these data. Although the interaction of Percentage of Reinforcement × Degree of Learning was not statistically significant, further tests of linear trend clarify the statistical relationships among the means. For the 100% reinforcement groups a test for linearity as a function of degree of learning was not statistically significant, $F = .69$. For the 50% reinforcement group this same test indicated a linear upward trend, $F\ (1, 144) = 6.62$, $p < .05$, suggesting that after 50% reinforcement there obtains a direct relationship between degree of learning and trials to criterion in reversal.

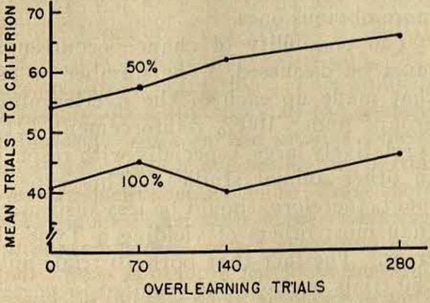


FIG. 1. Mean trials to criterion in reversal learning as a function of degree of learning and percentage reinforcement. ($N = 21$ for each point.)

TABLE 1

ANALYSIS OF VARIANCE FOR TOTAL TRIALS TO CRITERION DURING REVERSAL LEARNING

Source	*df*	*F*
Percentage of reinforcement (P)	1	46.62**
Degree of learning (D)	3	2.05
Level (L)	2	1.01
P × D	3	.86
P × L	2	4.51*
D × L	6	1.66
P × D × L	6	.48
Pooled error (*MS*)	144	(247.0436)
Total	167	

$^{*}p < .05$.
$^{**}p < .01$.

The means for the 100% and 50% reinforcement groups were 43.2 and 59.8, respectively. This significant difference indicates faster reversal following 100% than following partial reinforcement. The means for the three acquisition levels were 48.8, 52.8, and 52.2 for the fast, intermediate, and slow *S*s, respectively. This variation was not statistically significant. After 100% reinforcement the three acquisition levels had means of 38.3, 42.1, and 49.7, while after 50% reinforcement the means were 59.3, 63.5, and 54.8. This interaction was significant at the 1%

level of confidence, and indicates that after 100% reinforcement there is a direct relationship between speed of learning in acquisition and speed of reversal learning, while after 50% reinforcement, slow learners are faster reversers.

No other terms in the analysis of variance were statistically significant.

Errors.—The data were also analyzed according to the number of errors made during reversal learning. The results were essentially the same as those for trials to criterion and the relationship among the means for the four overlearning groups following 50% and 100% reinforcement was very much similar to the picture presented by trials to criterion. Analysis for linearity showed a significant increase in errors with an increase in degree of learning, $F\,(1, 96) = 8.30$, $p < .01$ for the 50% reinforcement condition. No such relationship was found in the 100% reinforcement group ($F = .01$).

DISCUSSION

The results showed that the 50% reinforcement groups were inferior to the 100% groups in reversal learning. This is consistent with the results of Wike (1953); Grosslight, Hall, and Scott (1954); Grosslight and Radlow (1956, 1957); and Kendler and Lachman (1958). That such a relationship should exist is not too surprising considering the similarity between reversal learning and simple extinction. The fact that partial reinforcement leads to greater resistance to extinction is now well documented (e.g., Lewis, 1960). It is very likely that the results found in reversal learning after partial reinforcement are due to increased resistance to extinction of responses to the positive acquisition stimulus. Partial evidence for this statement comes from an analysis of perseveration scores of the present data, where perseveration is defined as the number of trials to the first correct response in reversal. The 100% group took a mean of 12.8 trials, while the 50% group took 24.7 trials on the average, $F(1, 144) = 51.00$, $p < .01$.

In this study the overlearning effect that has been so often reported for reversal learning when the acquisition reinforcement was 100% was not found. No guess can be made at the moment as to the variables responsible for the discrepancy between the present results and those of others. A survey of variables that might at first hand seem to be operating here, shows that the conditions in the present experiment were in one combination or another present in other experiments. Thus, for example, a T maze was used by both Capaldi and Stevenson (1957) and by Brookshire et al. (1961); a brightness discrimination was learned in experiments by Reid (1953), Pubols (1956), Capaldi and Stevenson (1957), Birch et al. (1960), and by D'Amato and Jagoda (1961). In all rat reversal learning studies so far presented, at least 1 day intervened between acquisition or overlearning and reversal; white was the positive stimulus during acquisition in three studies (Birch et al., 1960; Capaldi & Stevenson, 1957; Pubols, 1956); Sprague-Dawley animals were used in two studies (Brookshire et al., 1961; Pubols, 1956); the intertrial intervals have varied from 20 sec. (Capaldi & Stevenson, 1957) to 6 min. (Pubols, 1956) with Reid's (1953) being exactly that of the present. This survey does not exhaust all the conditions that might be relevant, but it does include the more obvious ones.

The possibility of chance occurrence must be discussed. The number of *S*s that made up each of the overlearning groups under 100% reinforcement (21) is relatively large, especially with respect to other animal studies. This experiment, therefore, should be less suspected than most others of yielding a Type II error. The fact that both 70 trials and 280 trials overlearning resulted in *poorer* performance than did 0 trials overlearning reduces even more the likelihood of a Type II error. Furthermore, it must be mentioned that several recent experiments have failed to replicate the overlearning effect (D'Amato & Jagoda, 1962; Hill, Spear, & Clayton, 1962).

The results of this experiment do show that when acquisition is under 50% reinforcement, there is an inverse relationship between the degree of overlearning and ease of reversal learning. These results are consistent with those found in the card sorting situation (Erlebacher & Archer, 1961) and show that the relationship found under 50% reinforcement is not the same as that found under 100% reinforcement. The fact that differences exist between 50% and 100% reinforcement tends to indicate that the discrimination hypothesis, as defined in the introduction to this paper, does not apply to *both* the partial reinforcement effect and the overlearning effect in reversal learning. This should be true, since if the same hypothesis explained both effects, either effect should hold under all levels of the other. That is, the partial reinforcement effect should occur under all overlearning conditions and the overlearning effect should hold under all percentages of reinforcement.

Although the results of Erlebacher and Archer (1961) were consistent with the present results, they interpreted their findings as being consistent with the discrimination hypothesis. They reasoned that while learning under 50% reinforcement, with overlearning, *S*s learn a greater number of different sequences of nonreinforced responses that are not negatory in nature. There should therefore be a greater similarity between acquisition and extinction, leading to slower extinction. This analysis probably does not apply to the present experiment since in the present experiment a limitation was set on the partial reinforcement sequence so that there were never more than three reinforced or nonreinforced correct responses in a row. The *S*s even in the 0-overlearning condition therefore had ample experience with all types of sequences of nonreinforced trials.

Since the discrimination hypothesis cannot account for both the partial reinforcement effect and the overlearning effect, then if it is correct for one or the other of the two effects, a two- or many-factor hypothesis must account for the present results. It is, of course, possible that the discrimination hypothesis accounts for neither effect.

One more result need be discussed. This is the result that after 100% reinforcement, there was a direct relationship between acquisition speed and speed of reversal learning whereas after 50% reinforcement, the slow learners reversed the fastest. This result is probably due to the nature of the reversal learning task in that this task involves both the extinction of the formerly correct response and the learning of the new correct response. It can be assumed that there are reliable individual differences in learning such that *S*s who are slow at learning the first correct response are also slow at learning a new response. On the other hand, there is an inverse relationship between acquisition speed and speed of extinction (Hunter, 1935; Youtz, 1938). The reversal learning situation must therefore be governed by an interaction between the two factors. Since after 100% reinforcement extinction is relatively quick, it may be assumed that the differences in learning abilities is the more important factor, leading to the prediction that there be a direct relationship between acquisition and reversal performance. After 50% reinforcement, the extinction factor may be the more important, leading to the present results.

REFERENCES

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

BROOKSHIRE, K. H., WARREN, J. M., & BALL, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, **54**, 98–102.

BRUNER, J. S., MANDLER, J. M., O'DOWD, D., & WALLACH, M. A. The role of overlearning and drive level in reversal learning. *J. comp. physiol. Psychol.*, 1958, **51**, 607–613.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

D'AMATO, M. R., & JAGODA, H. Analysis of the role of overlearning in discrimination reversal. *J. exp. Psychol.*, 1961, **61**, 45–60.

D'AMATO, M. R., & JAGODA, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, **64**, 117–122.

ERLEBACHER, A. Reversal learning in rats as a function of percentage reinforcement and degree of overlearning. Unpublished doctoral dissertation, University of Wisconsin, 1961.

ERLEBACHER, A., & ARCHER, E. J. Perseveration as a function of degree of learning and percentage reinforcement in card sorting. *J. exp. Psychol.*, 1961, **62**, 510–517.

GELLERMAN, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42**, 207–208.

GRANT, D. A., & SCHIPPER, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

GROSSLIGHT, J. H., HALL, J. F., & SCOTT, W. Reinforcement schedules in habit reversal: A confirmation. *J. exp. Psychol.*, 1954, **48**, 173–174.

GROSSLIGHT, J. H., & RADLOW, R. Patterning effect of the nonreinforcement-reinforcement sequence in a discrimination situation. *J. comp. physiol. Psychol.*, 1956, **49**, 542–546.

GROSSLIGHT, J. H., & RADLOW, R. Patterning effect of the nonreinforcement-reinforcement sequence involving a single nonreinforced trial. *J. comp. physiol. Psychol.*, 1957, **50**, 23–25.

HILL, W. F., SPEAR, N. E., & CLAYTON, K. N. T maze reversal learning after several different overtraining procedures. *J. exp. Psychol.*, 1962, **64**, 533–540.

HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

HUNTER, W. S. Conditioning and extinction in the rat. *Brit. J. Psychol.*, 1935, **6**, 135–148.

KENDLER, H. H., & LACHMAN, R. Habit reversal as a function of schedule of reinforcement and drive strength. *J. exp. Psychol.*, 1958, **55**, 584–591.

LEWIS, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57**, 1–28.

MURILLO, N. R., & CAPALDI, E. J. The role of overlearning trials in determining resistance to extinction. *J. exp. Psychol.*, 1961, **61**, 345–349.

PUBOLS, B. H. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of non-continuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

SENKO, M. G., CHAMP, R. A., & CAPALDI, E. J. Supplementary report: Resistance to extinction of a verbal response as a function of the number of acquisition trials. *J. exp. Psychol.*, 1961, **61**, 350–351.

STEVENSON, H. W., & MOUSHEGIAN, G. The effect of instructions and degree of training on shifts of discriminative responses. *Amer. J. Psychol.*, 1956, **69**, 281–284.

STIMMEL, D. T., & NORTH, A. J. Effects of extended acquisition training on subsequent extinction of a running habit. *Amer. Psychologist*, 1959, **14**, 420.

WIKE, E. L. Extinction of a partially and continuously reinforced response with and without a reward alternative. *J. exp. Psychol.*, 1953, **46**, 255–256.

YOUTZ, R. E. P. Reinforcement, extinction, and spontaneous recovery in a non-Pavlovian reaction. *J. exp. Psychol.*, 1938, **22**, 305–318.

(Received August 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 91–94

HUMAN REINFORCEMENT: EXPERIMENTER AND SUBJECT CONTROLLED [1]

ALBERT R. MARSTON AND FREDERICK H. KANFER

University of Wisconsin *University of Oregon Medical School*

The concept of self-reinforcement was examined as an example of control of human behavior without direct, external reinforcement. Ss at 3 levels of incentive were given a learning period on a verbal discrimination task, with a light as reinforcer. In Phase II, they received 1 of 3 reinforcing procedures: SR (in which S was given complete control of administering reinforcing stimuli), continued acquisition, and extinction. The 3 procedures differed significantly in frequency of correct responses during Phase II, with the SR group lying between acquisition and extinction. The highest incentive group showed the greatest accuracy in taking available reinforcement; incentive did not affect total frequency of reinforcements taken.

A recently completed study (Kanfer & Marston, 1963) investigated learning under conditions in which *S* received reinforcement vicariously by observing administration to others. The present experiment deals with another example of control of behavior without direct, external reinforcement. Skinner's (1953) concept of *self-reinforcement* attempts to handle the observation that much of our learned behavior is maintained without the continued external administration of even highly generalized reinforcers. By relying on earlier experience, the individual may serve as his own source of reinforcing stimuli either by manipulating the conditions for obtaining external reinforcement or by administering social reinforcers through subvocal responses. The procedure used in the current experiment brings self-reinforcement into the laboratory in the observable form of a motor response that secures an unlimited, available supply of reinforcing stimuli.

An earlier study (Kanfer, Bradley, & Marston, 1962) involved an initial attempt to relate self-reinforcement to amount of learning. It was found that an increase in number of learning trials led to higher frequency of self-administered reinforcement (SRs) following correct responses. One purpose of the current investigations is to compare the relative effect of self-reinforcement and the usual acquisition and extinction procedures upon strength of a learned response. A second purpose of the present study is to compare three levels of incentive in their effects upon self-reinforcement.

METHOD

Subjects.—Sixty-two male undergraduate students participated in the experiment.

Apparatus.—The *S* was seated at a table facing a 4 × 5 ft. black screen which separated him from *E*. Stimulus cards were presented manually in a window cut into the screen at eye level. A green light, placed just below the window, served as a reinforcing stimulus. The *E* activated the light for .3 sec. by pressing a microswitch. Groups receiving white poker chips as reinforcing stimuli received them in a cup placed beneath a slot cut at the base of the board.

Procedure.—In Phase I, *S* was shown a series of 10 cards, each containing a different set of four nonsense syllables of medium association value typed in a horizontal line.

---

[1] This research was supported in part by Research Grant M-2027 (C4) from the National Institute of Mental Health, United States Public Health Service, F. H. Kanfer, Principal Investigator. The authors wish to acknowledge the assistance of Marcia Bradley.

Instructions indicated that *S* was to spell one syllable on each trial and that the reinforcing stimulus (light) would indicate a correct choice. The position of the correct syllable was varied randomly over the 10 cards and each of the four syllable positions was designated as correct with equal frequency. The 10 cards were shuffled and used again on subsequent blocks; the same syllables were correct on each presentation. Learning continued until *S* reached a criterion of 6/10 correct trials in a block.

Three levels of incentive were used. *Low incentive* consisted of the green light alone. *Medium incentive* was represented by white poker chips with each light flash. For *high incentive* a prize package was given for each 30 chips. The prizes, opened at the end of the session, were displayed before *S*s in this group and included such small items as pens and pocket knives and the possibility of a chance on a drawing for the two larger prizes, a $10.00 gift certificate or dinner for two at a local restaurant. Using 30 chips in exchange for a prize allowed approximately four prizes per *S*.

The comparison of *S*-controlled reinforcement, externally controlled reinforcement (continued acquisition), and extinction was made during Phase II after each *S* reached the criterion in Phase I. In the acquisition group (AC), there was no break in the procedure and *E* continued to reinforce correct responses for 10 additional blocks of 10 trials each. In the extinction (EXT) group, the 10 additional blocks were given without reinforcement. The self-reinforcement (SR) group was told that the procedure would continue as before but that *E* would no longer operate the switch controlling the reinforcing stimuli. The *S* was given the switch and instructed to press it whenever he was fairly confident he was correct.

Results were analyzed in two parts: (*a*) the effect of the reinforcement procedure and incentive on variations in correct responses during Phase II and (*b*) the effect of incentive on frequency of SRs and on accuracy of SRs (i.e. proportion of incorrect SRs). Each cell of the complete factorial design for Part *a* of the analysis contained 4 *S*s. In Part *b*, comparing the three incentive levels within the SR procedure, the number of *S*s in each of the cells was increased to 13.

RESULTS AND DISCUSSION

Preliminary to the analysis of Phase II, the treatment groups were compared on measures of the acquisition given all *S*s in Phase I. Incentive conditions did not affect number of trials to criterion, number of reinforcements to criterion, or frequency of correct responses on the final block of Phase I (all *F*s < 1.0). Although they did not differ in treatment until Phase II, the AC, EXT, and SR groups were also compared on these three measures and found not to differ significantly (*F*s = 3.06, 1.87, and 1.77, respectively).

TABLE 1

ANALYSIS OF VARIANCE ON CORRECT RESPONSES DURING PHASE II

Source	*df*	*MS*	*F*
Reinforcement condition (A)	2	363.27	10.93***
Incentive level (B)	2	2.04	0.06
A × B	4	21.73	0.65
*S*s in groups (*S*s)	27	33.21	
Blocks (C)	9	4.65	3.07**
A × C	18	4.79	3.17**
B × C	18	1.10	0.72
A × B × C	36	1.97	1.30
C × *S*s	243	1.51	

** $p < .01$.
*** $p < .001$.

The relative effect of the three reinforcing procedures in Phase II was examined by an analysis of frequency of correct responses during the 10 blocks of trials given in this period (see Table 1).

Incentive conditions did not affect correct responses, but the reinforcement procedure and its interaction with blocks were significant ($F = 10.93$, $p < .001$ and $F = 3.17$, $p < .01$, respectively). The AC group increased in response strength during Phase II and was the highest of the three, with the SR group next, and EXT lowest (see Table 2). Block effects tests in each of the three groups showed that AC increased significantly over blocks, $F\ (9, 243) = 7.33$, $p < .001$, while EXT decreased sig-

TABLE 2

Mean Frequency of Correct Responses during Phase II for the Three Reinforcing Procedures

Group	Blocks									
	1	2	3	4	5	6	7	8	9	10
AC	5.8	5.8	7.1	7.4	7.8	7.4	8.2	8.0	8.5	8.4
SR	5.2	5.2	4.7	5.3	5.4	4.8	4.8	5.3	5.3	5.1
EXT	4.9	4.0	4.1	2.9	4.3	4.8	3.8	3.9	3.9	3.9

nificantly, $F\ (9, 243) = 2.46, p < .01$, and SR did not change ($F < 1.0$). A Tukey gap test indicated that all three group means differed from one another significantly at the .05 level (means for 10 blocks = 74.42 in Group AC, 51.00 in Group SR, and 42.92 in Group EXT; critical difference = 4.77).

The results of these analyses indicate that the effect of self-reinforcement on response strength lies between that of continued acquisition and the usual extinction procedure. While self-reinforcement did not increase response strength in this situation, neither did it decrease frequency of correct responses as did extinction.

The possibility that instructions might serve as a source of response bias led to a procedural problem in determining the timing of instructions for the self-reinforcement procedure. Although it would have been possible to give the SR group their Phase II instructions earlier in the session, doing so would have raised additional problems concerning *S*s' differential recalls and would not have eliminated other critical stimulus cues differentiating Phase II from Phase I of the procedure. That these instructions alone did not appear to account for the findings is indicated by the fact that the greatest differences between the SR and EXT groups occurred not on the first block of Phase II but on subsequent blocks. A *t* test modified for repeated comparisons of pairs of means (Lindquist, 1953) showed that these groups differed significantly on Blocks 2, 4, 5, 8, 9, and 10.

Table 3 presents a summary of the analyses of variance comparing the three incentive groups on two meas-

TABLE 3

Analyses of Variance of Total SR Frequencies and Arc-Sine Transformations of Proportion of Incorrect SRs to Total SRs

Source	*df*	*MS*		*F*	
		Total *f*	Propor. Incorrect	Total *f*	Propor. Incorrect
Levels of incentive (A)	2	7.40	8.13	1.0	3.26*
*S*s in groups (*S*s)	36	30.66	2.49		
Blocks (B)	9	6.39	0.15	4.66***	1.36
A × B	18	1.15	0.04	1.00	1.00
B × *S*s	324	1.37	0.11		

* $p < .05$.
*** $p < .001$.

TABLE 4

MEAN PROPORTION OF INCORRECT SRs TO TOTAL SRs DURING PHASE II FOR THREE INCENTIVE GROUPS

Groups	Blocks									
	1	2	3	4	5	6	7	8	9	10
Low motivation	.28	.27	.31	.29	.29	.36	.32	.31	.34	.38
Medium motivation	.31	.30	.35	.37	.36	.39	.37	.37	.39	.40
High motivation	.17	.19	.19	.19	.20	.19	.16	.18	.18	.19

ures of self-reinforcement: frequency of SRs and proportion of incorrect SRs to total SRs (converted to arc sines). Although frequency of SRs increased over blocks for all groups ($F = 4.66$, $p < .001$), there were no significant effects due to incentive level. The *S*s administered SRs to a response with a probability quite close to that with which the response was reinforced by *E* at the end of acquisition. The mean frequency of SRs on the first block of Phase II was 6/10, and the mean frequency of correct responses on the last block of Phase I was 6.5/10. Given relatively little instructional information directly relevant to the task, *S* appears to have handled the ambiguity by continuing as before, i.e. by maintaining the previous frequency of correct responses and approximating the final frequency of external reinforcement.

The analysis of mean proportion of incorrect SRs indicated that incentive groups differed significantly ($F = 3.26$, $p < .05$). The proportion of the total number of SRs that were administered following incorrect responses was smallest for the High incentive condition, with the Low and Medium incentive groups increasing in that order (see Table 4). High incentive produced greater caution, i.e. fewer incorrect responses were given SRs than in either of the other incentive conditions.

The results indicate that the level of incentive affected neither frequency of correct responses nor frequency of SRs but did affect the distribution of SRs to correct and incorrect responses. The fact that the smaller proportion of incorrect to total SRs in the High incentive group was not reflected in an increase in correct responses suggests that self-reinforcement operates somewhat differently than external reinforcement. Thus, the findings appear to show that distribution of SRs can be manipulated by such external conditions as incentive, independently of the learned response for which SR is given.

REFERENCES

KANFER, F. H., BRADLEY, M. M., & MARSTON, A. R. Self reinforcement as a function of degree of learning. *Psychol. Rep.*, 1962, **10**, 885–886.

KANFER, F. H., & MARSTON, A. R. Human reinforcement: Vicarious and direct. *J. exp. Psychol.*, 1963, **65**, 292–296.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

(Received August 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 95–99

EFFECT OF VARIOUS DISPLAY-CONTROL CONFIGURATIONS ON TRACKING WITH IDENTICAL AND DIFFERENT COORDINATE DYNAMICS

RUBE CHERNIKOFF AND MOIRA LEMAY

United States Naval Research Laboratory, Washington, D. C.

Position and acceleration dynamics were paired in a 2-coordinate tracking task so that both identical and different dynamics were used with the following display-control configurations: 1-dot display, 1-stick control; 1-dot display, 2-stick control; 2-dot display, 1-stick control; and 2-dot display, 2-stick control. The results showed a striking shift in the order of tracking proficiency for the various display-control configurations with the change from identical to different coordinate dynamics. When the same dynamics are used, the influence of displays predominates. However, when different dynamics are used in each coordinate, controls are the important determiners of accuracy.

In most conventional tracking systems the display consists of a single spot of light free to move over the face of a cathode-ray tube, and control is by means of a stick which can be positioned at any angle in one or two dimensions. As long as the tracking task involves only one coordinate, or two coordinates both of which use the same system dynamics, this single-dot, single-stick arrangement provides adequate control. However, Chernikoff, Duey, and Taylor (1960), using various pairings of control dynamics, found that error increased sharply as a function of the degree of difference in dynamics between the two coordinates. Conceivably, other methods of information display and tracking control could be found to provide superior performance in systems requiring different dynamics in each coordinate. A logical possibility for investigation is that of providing separate display information and separate controls for each coordinate. There has been little previous research done with such configurations. The most relevant information is provided in a study by Charipper (1959). He compared a single-element and a two-element display in combination with various types of controls, and found the single-element display to be superior with each of three types of controls.

The present investigation was concerned with determining the effects on tracking performance, both with different and with identical control dynamics in two coordinates when separate display information or separate controls are provided for each coordinate. Position and acceleration (zero and second order) dynamics were paired in two coordinates so that systems of both identical and different dynamics were tracked under the following display-control configurations:

1. One-dot, one-stick: This is the conventional two-coordinate compensatory tracking system, in which both the display dot and control stick are free to move in both coordinates.

2. One-dot, two-stick: In this condition the display dot was free to move in two coordinates, but one stick controlled only the deviations in the X coordinate while a second stick controlled error in the Y coordinate.

3. Two-dot, one-stick: One dot moved only along the X coordinate and a second dot moved only along the Y coordinate of the display. The

control was a single stick free to move in both X and Y.

4. Two-dot, two-stick: Both the display dots and control sticks are separated by coordinates. Thus, the dot moving along the X coordinate is controlled by a stick free to move only in X, and the dot moving along the Y coordinate can be simultaneously controlled by another stick restrained to move only along Y.

METHOD

Apparatus.—Figure 1 shows a block diagram of the tracking system. Selection of the type of control dynamics to be used in each coordinate is made by Switches S_x and S_y. When either switch is set on Contact P, position control is obtained in the corresponding coordinate, while setting a switch at Contact A produces acceleration control. Each switch can be set separately, so all combinations of dynamics in the X and Y coordinates can be obtained.

The course consisted of two sine waves with frequencies of 4.7 and 2.9 cycles/min. Each sine wave was fed into both coordinates, 90° apart. The amplitudes of the sine waves were inversely proportional to their frequencies, and sufficient to displace the tracking dot(s) over the middle 4 in. of the 5-in. short persistence cathode-ray tube (CRT) display. The resulting course was equal in the two coordinates. A compensatory tracking system was used. That is, variations in the course moved the dot(s) away from a center reference, and *S* made the appropriate control movements to return the dot(s) to center. Since a compensatory system displays the course as modified by *S*'s corrective movements, the course could not be learned. The control-display gain for both coordinates was such that a 10° displacement of the stick produced a tracking dot change on the CRT face of 1 in. with position control, and 1 in/sec^2 with acceleration control.

Tracking error, consisting of the difference between the course input and the system output, was scored separately for each coordinate by use of two identical electronic scoring integrators. The error for each coordinate was rectified and integrated over the scoring duration, and read from a voltmeter. Since the time of each tracking trial was held constant, integrated error is proportional to average error, and can be used to represent the average error for the trial.

An "arbitrary units" scale was used to present the average error scores. However, the ratio of any error score to the average absolute value of system input can be determined by multiplying the arbitrary units score by .000625. Thus, for example, a ratio of .025 is obtained for a score of 40 arbitrary units. In other words, this ratio indicates that the absolute average error was only 2.5% of the absolute average course value.

Procedure.—The three pairings of dynamics (position-position, position-acceleration, and acceleration-acceleration), combined with the four display-control combinations, gave rise to 12 experimental conditions. These were presented to each *S* in a different random order during each experimental session. In addition, a counterbalanced design was used for the pairings of different dynamics so that both position and acceleration were used on the X and Y coordinates an equal number of times. The integrated error score for each coordinate was recorded separately. Scores in one coordinate for a given display-control condition with different dynamics were com-

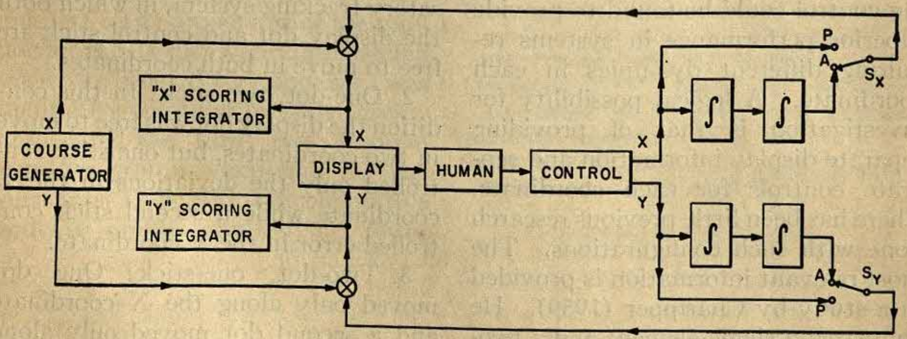


FIG. 1. Block diagram of the tracking system.

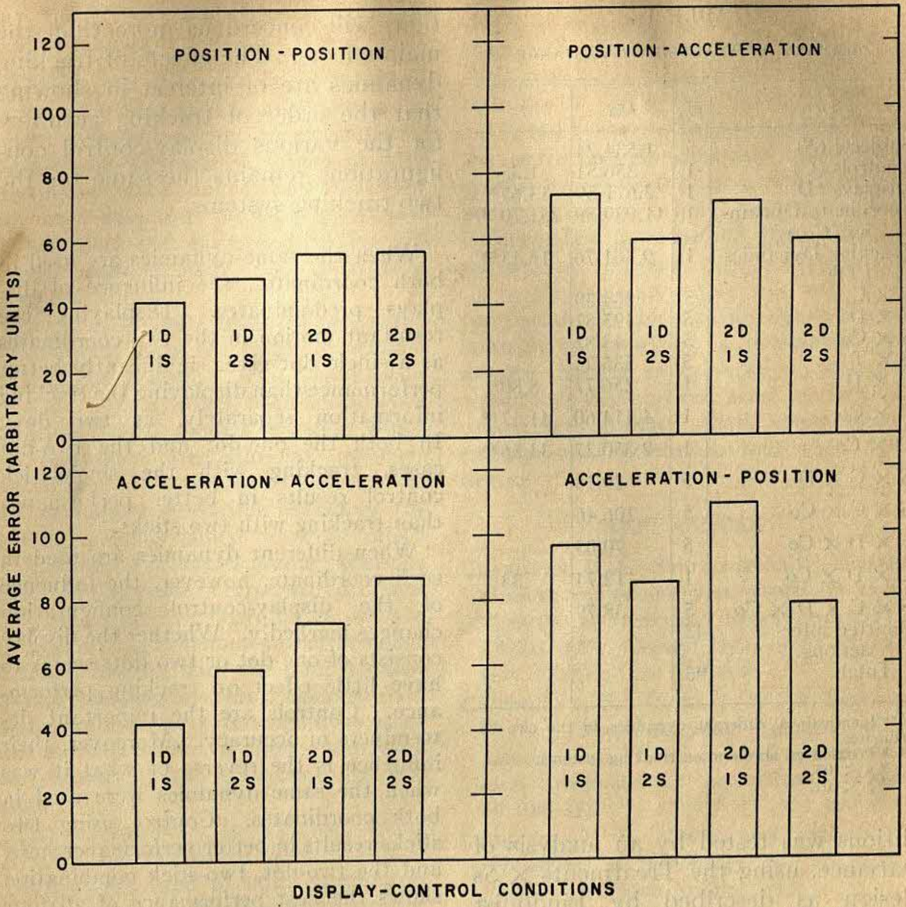


FIG. 2. Average error for the various display-control conditions with identical and different coordinate dynamics. (The bars are labeled with the display dot and control stick configurations they represent.)

pared with the scores in the corresponding coordinate for the display-control condition with identical dynamics. When the two-stick control was used, each coordinate was tracked with the preferred and nonpreferred hands an equal number of times.

Six naval enlisted men served as *S*s. Each man tracked the 12 conditions for 24 sessions. A session consisted of three 1-min. trials on each of the conditions, the integrated error score for the last 45 sec. being recorded as the score for that trial. The mean for the three trials was then taken as *S*'s score for that session.

RESULTS AND DISCUSSION

The mean score for the last four sessions of tracking was obtained for each *S*, and a group mean was determined for each of the 12 experimental conditions. These group means are presented in the graph of Fig. 2. It should be noted that "position-acceleration" and "acceleration-position" represent the same condition, and that the bars in the upper right quadrant of Fig. 2 show the position scores, while those in the lower right quadrant depict the acceleration scores.

The significance of the observed differences in tracking performance due to the various experimental con-

TABLE 1

Summary of Analysis of Variance

Source	*df*	*MS*	*F*
Subjects (*S*)	5	1,534.24	
Controls (C)	1	356.51	1.35
Displays (D)	1	2,677.59	13.67*
Coordinate Dynamics (Co)[a]	1	11,029.59	251.70**
Tracking Dynamics (T)[b]	1	9,861.76	18.43**
S × C	5	264.39	
S × D	5	195.87	
S × Co	5	43.82	
S × T	5	535.24	
C × D	1	256.77	8.80*
C × Co	1	4,414.60	41.47**
D × Co	1	2,350.27	33.56**
S × C × D	5	29.18	
S × C × Co	5	106.46	
S × D × Co	5	70.03	
C × D × Co	1	12.74	.33
S × C × D × Co	5	38.79	
Omitted interactions	42		
Total	95		

[a] Identical vs. different dynamics in the two coordinates.
[b] Position vs. acceleration tracking systems.
* $p < .05$.
** $p < .01$.

ditions was tested by an analysis of variance, using the Treatments × *S*s design as described by Lindquist (1953).

It is evident from Fig. 2 that the influence of the display-control configuration varies considerably depending on whether or not both coordinates have the same dynamics. This result is supported by the analysis, presented in Table 1, which shows significant interactions both for Controls × Coordinate Dynamics and Displays × Coordinate Dynamics. Thus, the effects of controls, and of displays, are different under different conditions of coordinate dynamics. Since the differences engendered in the display-control configurations by the use of different dynamics in each coordinate are of primary interest, the interactions will concern us more than the main effects. The effects of tracking dynamics are of interest in showing that the order of tracking accuracy for the various display-control configurations remains the same for the two tracking systems.

When the same dynamics are used in both coordinates, the influence of displays predominates. Displaying the resultant motion in the two coordinates as a single dot gives significantly better performance than displaying the tracking information separately, as two dots. In both the one-dot and the two-dot cases, tracking with the single-stick control results in better performance than tracking with two sticks.

When different dynamics are used in each coordinate, however, the influence of the display-control configuration changes markedly. Whether the display consists of one dot or two dots seems to have little effect on tracking performance. Controls are the important determiners of accuracy. Moreover, their influence is the reverse of what it was when the same dynamics were used in both coordinates. Control using two sticks results in better performance here, and the two-dot, two-stick combination shows the best performance of all four display-control conditions with acceleration-position dynamics.

These results are borne out by the fact that while the main effect of controls is not significant, the interaction of controls and coordinate dynamics is significant. When identical and different coordinate dynamics are combined in measuring the main effect of controls, the control differences compensate for one another, and significant effects are masked. However, when the effect of controls is computed separately for different conditions of coordinate dynamics, as in the interaction term, it is highly significant. Thus, tracking performance may be improved by use of the appropriate control configuration for different situations of coordinate dynamics.

From the above results it can be seen

that tracking performance with several display-control configurations is significantly affected by whether or not the dynamics are the same in both coordinates. The increased error introduced in a conventional one-dot, one-stick tracking task by different dynamics in each coordinate may be reduced by displaying the error separately in each coordinate by means of two dots, and having the tracker control each dot separately, using two sticks.

The most striking result of this study is the alteration of the effects of display-control configuration with the change from a system in which both coordinates have identical dynamics to one in which the dynamics are different in each coordinate. This change in effects gives rise to some speculation regarding the operator's performance on what might be considered a multichannel tracking task. The increase in error with different dynamics in the two coordinates, and the subsequent improvement in performance in that situation with the dual display-control configuration, are consistent with the hypothesis that the operator must time-share in order to track more than one channel if the dynamics are dissimilar.

The operator is, in a sense, being "forced" to time-share when identical dynamics are used in both coordinates by separating the displays and controls. This is so, since he can no longer treat the identical coordinate situation as a single dimension, and make a single control response along some diagonal axis which is appropriate for both coordinates. When dynamics are different in the two coordinates, he must treat the error which develops along one coordinate separately from what is happening along the other. This separation is necessary because of the different pattern of response movements called for by the dynamics associated with each coordinate. When separate displays and controls are provided, it is plausible to assume that the experienced tracker can simultaneously null out error in the two coordinates, rather than sequentially control one coordinate and then the other.

REFERENCES

CHARIPPER, B. A. Ship control VIII. Single-element vs. two-element display in two-dimensional tracking. Report No. SPD 59-003, 1959, General Dynamics Corporation, Electric Boat Division.

CHERNIKOFF, R., DUEY, J. W., & TAYLOR, F. V. Two-dimensional tracking with identical and different control dynamics in each coordinate. *J. exp. Psychol.*, 1960, 60, 318–322.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

(Received August 27, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 100–106

STIMULUS DURATIONS AND STIMULUS CHARACTERISTICS IN PAIRED-ASSOCIATES LEARNING[1]

CALVIN F. NODINE[2]

University of Massachusetts

Acquisition of paired-associates units was investigated as functions of orthogonal combinations of duration of stimulus members alone and of duration of stimulus and response members, and of similarity of stimulus and of response members. Responses recalled correctly were direct functions of durations of stimulus members, durations of stimulus and response members, meaningfulness of stimulus and response members; they were inverse functions of similarity of stimulus and response members. These effects were more potent on the response than on the stimulus side. Meaningfulness and similarity are conceived as acting through stable, discriminable recognition responses to stimulus and response members. Lengthening durations presumably provides greater opportunity for occurrence and rehearsal of such recognition responses.

Experiments concerned with effects of variations in temporal aspects of presentation of paired-associates (PA) units have been infrequent and limited in scope. The primary objective of the present investigation was the systematic exploration of two temporal parameters associated with the anticipation technique of presentation: exposure duration of the stimulus member of the PA unit alone (St duration), and exposure duration of the stimulus and response member paired (St-R duration). Of secondary interest in the present experiment was information concerning relative rates of acquisition of PA units representing different combinations of meaningfulness (M) and similarity (Sim) of stimulus and response units as functions of combinations of St and St-R durations.

Previous experiments which have varied St and St-R durations jointly, but in confounded fashion, include Hovland (1949) using 1:1 sec. and 2:2 sec. St, St-R durations; Goss, Morgan, and Golin (1959) using 2:2 sec. and 3:3 sec. durations; and Wilcoxon, Wilson, and Wise (1961) using 1:1, 2:2, and 4:4 sec. durations. These studies indicate a strong direct relationship between exposure durations and acquisition rate. Total learning time was, however, invariant.

Bugelski (1962) varied St-R durations from 2 to 15 sec. independently of a fixed 2-sec. St duration. Trials to criterion, but not total learning time, were found inversely related to St-R durations. Bugelski's experiment together with those reported earlier demonstrate a need for parametric assessment of PA learning as a function of St, St-R durations.

Acquisition rate has been found directly related to M and inversely related to Sim of stimulus and response units with M more potent on the response side and the converse true of Sim (e.g., Goss & Nodine, 1963; Goss, Nodine, Gregory, Taub, & Kennedy, 1962). Because these

---

[1] This paper is based upon a PhD dissertation submitted to the Department of Psychology of the University of Massachusetts. The author is greatly indebted to A. E. Goss under whose direction this investigation was conducted. The analysis of variance was processed, in part, at the Computation Center, Massachusetts Institute of Technology.

[2] Now at American University, Washington, D. C.

TABLE 1

List of PA Units Representing Combinations of Low (L) or High (H) Similarity (Sim) and Meaningfulness (M) of Stimulus and Response Members

Stimulus Members		Response Members				Mean M
Sim	M	LL	LH	HL	HH	
H	H	CUD(95)[a] ZAS(25)	DUC(90) HOB(90)	CUS(93) VEJ(22)	SUC(90) RAK(85)	92.00
H	L	ZOV(15) GIC(19)	FOV(27) CIT(93)	VOZ(27) QEJ(6)	VOF(10) WAK(87)	19.75
L	H	BAH(94) YED(26)	GEL(89) TUF(89)	PIX(94) JEV(21)	MIN(94) KAW(80)	92.75
L	L	NAJ(18) FEP(14)	KUG(29) SOL(94)	YIB(16) JEQ(14)	WOY(22) KAR(92)	21.25
Mean M		21.00	91.50	15.75	86.00	

Note.—See text for interpretation of table.
[a] Archer 1960 M values.

stimulus characteristics might be expected to interact with exposure durations, a subsidiary objective of the present experiment was to obtain data on the nature of these interactions.

Lengthening St and St-R durations could be regarded as providing increased opportunity for *S*s to form stable, differential recognition responses to stimulus and response members, hastening integration and increasing their availability for S-R association. It was therefore expected that as durations lengthen, any initial advantage due to specific combinations of stimulus characteristics should disappear.

In the present experiment, St and St-R durations of ½, 1, 2, and 4 sec. were combined orthogonally to yield parametric information on the course of acquisition as functions of different combinations of St, St-R durations. Sixteen consonant-vowel-consonant (CVC) trigrams were incorporated in a mixed list in which each unit represented a different combination of high or low M and high or low Sim of both stimulus and response members.

Method

Subjects.—Data are reported on 160 undergraduate *S*s drawn from the introductory psychology courses at the University of Massachusetts. A total of 10 *S*s, 5 males and 5 females, were assigned at random to each of the 16 combinations of St, St-R durations.

Apparatus.—The apparatus components consisted of a Stowe memory drum and an electronic voice key, microphone, and timer. Rotation of the memory drum and activation of the shutter mechanism used to expose the response member were electronically controlled in a manner which permitted independent variation of St and St-R durations.

The voice key was used to obtain latencies of *S*s' responses during St durations on recall-test trials. The timer was started simultaneously with exposure of each stimulus member and stopped either by *S*'s verbal response or automatically at termination of exposure of that stimulus member.

List.—The list was made up of 16 pairs of CVC trigrams. Each pair represented 1 of the 16 possible combinations of high or low Sim and high or low M of both stimulus and response members. Table 1 shows these 16 pairs. Particular combinations of Sim and M of stimulus members are indicated by a letter code in which the first L or H indicates low or high Sim and the second L or H indicates low or high M. Combinations of Sim and M of response members are coded in the same manner. Thus for the pair CUD ZAS, the stimulus member CUD is coded HH in the first row indicating CUD is of high Sim and of high M (95% Archer M). The response member ZAS is coded LL in the first column indicating ZAS is of low Sim and low M (25% Archer M). Other combinations are coded accordingly.

The mean values of high and low M for stimulus members are shown in the right column and those for high and low M of response members in the bottom row. The mean values for high M were close to the maximum attainable while those for low M were close to the minimum attainable within constraints imposed by the requirement that stimulus and response members also be of high or low Sim.

Similarity was manipulated in terms of number of common letters among stimulus members or response members. Within each of the two sets of four stimulus members of high Sim (Rows 1 and 2) and within each of the two sets of four response members of high Sim (Columns 3 and 4), each member had a counterpart in reverse with which it had three letters in common. Each member of a set also had two letters in common with the remaining two members of that set. Within each of the two sets of four stimulus members of low Sim (Rows 3 and 4) and within each of the two sets of four response members of low Sim (Columns 1 and 2), no member had any letter, with the exception of vowels, in common with those of any of the other three members of that set. Overlap of vowels and consonants of the 16 stimulus members with the 16 response members was minimal.

Serial learning was minimized by the use of five semirandom orders of the pairs for acquisition trials and four different orders for recall-test trials. All trigrams were typed in pica capitals.

Procedure.—The same list was acquired under all combinations of St and of St-R durations of ½, 1, 2, and 4 sec. with a constant 8-sec. intertrial interval. Following instructions which described the task, *S*s were administered four five-trial acquisition blocks by the conventional anticipation technique. Only correct anticipations which occurred during the St duration were recorded.

Since *S*s could not make anticipations in combinations containing ½-sec. St durations, it was necessary to employ interpolated test trials in which St durations were lengthened in order to provide information on the course of learning for these durations. Therefore each acquisition block was followed by a recall-test trial during which only stimulus members for each PA unit were presented. Regardless of the combination of St, St-R durations for acquisition, a 4-sec. St duration was maintained for recall-test trials. On these trials *S*s spoke into a microphone so that response latencies could be obtained. Only correct responses which occurred within the 4-sec. St duration were recorded. Incorrect responses or failures to respond within the 4-sec. St duration were arbitrarily assigned response latencies of 4 sec.

In order to have anticipation trials immediately before and immediately after each recall-test trial, the last recall trial was followed by one additional anticipation trial. Thus, in all, *S*s were given 21 anticipation trials and four systematically interpolated recall-test trials.

RESULTS

The three response measures were frequency of correct responses during each recall trial, mean number of correct responses for the anticipation trial immediately preceding and following each recall trial, and response latency during each recall trial. Figure 1 shows the means for each of these measures for all 16 PA units combined under each duration combination for four recall and four pairs of adjacent anticipation trials.

Correlations between corresponding duration combinations for recall frequencies and those for anticipation frequencies and latencies revealed certain relationships between response measures. The *r* for the relationship between means of recall and anticipation frequencies for the ½-sec. St duration was .00 since no *S* was able to anticipate in ½ sec. Difficulty in anticipating in less than 1 sec. was also evidenced by the *r* of .85 for the 1-sec. St duration in contrast to *r*'s of .96 and .97 for the 2- and 4-sec. St durations.

Latency means which included latencies of correct and incorrect responses could be viewed essentially as means of weighted correct recall frequencies. This may have accounted, in part, for the high *r*'s ranging from −.98 to −.99 for relationships between means of recall frequencies and recall latencies for each of the St durations disregarding St-R durations.

Because of these high correlations between response measures, only recall frequencies summed over Trials 1 and 2 (Trial Block 1) and Trials 3 and 4 (Trial Block 2) were treated by analysis of variance. The analysis of variance summary table contains 7

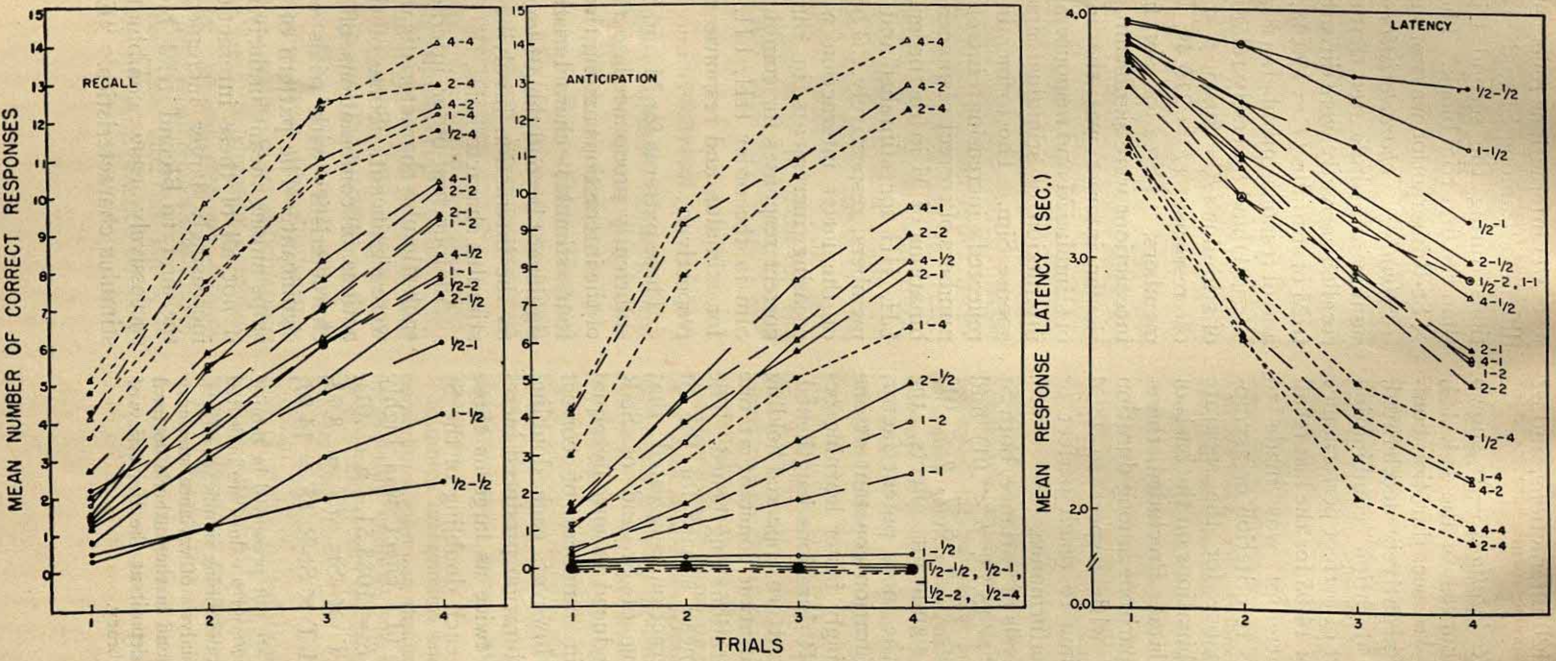


FIG. 1. Means of numbers of correct responses for recall and anticipation trials and means of response latencies for recall trials.

main effects, 120 interactions, and 32 error terms.[3]

St and St-R durations.—As shown in Fig. 1, the curves for the 4:4, 2:4, 4:2, 1:4, and $\frac{1}{2}$:4 sec. duration combinations are negatively accelerated. The longest duration combination (4:4) fell short of perfect performance with a mean of 14.05 for the last recall trial. No attempt was made to estimate trials to criterion or asymptotic performance for the separate duration combinations on the basis of the present data. Precluded, therefore, are comparisons among duration combinations with respect to total amount of time to reach perfect or asymptotic performance.

In the analysis of variance, both St durations ($F = 15.73$, $p < .01$) and St-R durations ($F = 56.57$, $p < .01$) had highly significant effects, the latter being the more potent factor. Holding St duration constant at one level, increasingly faster learning occurred as St-R durations lengthened. An identical pattern appeared holding each St-R duration constant at one level as St durations lengthened. The very small mean square for the St × St-R interaction suggested that each doubling of St and of St-R durations produced essentially equal increments in numbers of correct responses. However, each doubling of the St-R duration produced increments about twice as large as those produced by each doubling of the St duration.

The trials (T) effect was highly significant ($F = 1088.51$, $p < .01$). Significant T × St ($F = 8.73$, $p < .01$), and T × St-R ($F = 14.97$, $p < .01$) interactions reflected greater differences among St and St-R durations during the second trial block than the first trial block. The T × St × St-R interaction was not significant.

Stimulus characteristics.—Highly significant *F*s indicated that responses recalled correctly were directly related to M of stimulus ($F = 190.48$, $p < .01$) and of response members ($F = 510.78$, $p < .01$) and inversely related to Sim of stimulus ($F = 80.45$, $p < .01$) and of response ($F = 210.49$, $p < .01$) members. Three of six first-order interactions were significant. Of particular interest was the interaction of stimulus M and response M and the interaction of stimulus Sim and response Sim. The former interaction reflected increasing increments in numbers of correct responses for combinations of M in the order LL, HL, LH, HH for stimulus and response members, respectively. The pattern of the latter interaction was of increasing increments in numbers of correct responses for combinations of Sim in the order HH, LH, HL, LL for stimulus and response members, respectively.

These patterns for M and Sim were sufficiently pronounced so that three of the interactions among three of the four stimulus characteristics were significant beyond the .01 level as well as the interaction involving all four stimulus characteristics.

All four first-order interactions of each stimulus characteristic with trials were significant. Some of the interactions of combinations of stimulus characteristics with trials were also significant. The pattern of relationships among means underlying most, if not all of these interactions, was increasingly larger increments from the first to second trial blocks for successively easy combinations of stimulus characteristics. "Easy" referred to here and subsequently is defined empirically, i.e., in terms of number of correct responses obtained, as described above.

---

[3] The effects are not presented in tabular form, however, *F* and *p* values are indicated in the text when describing significant effects. The complete analysis of variance summary table may be found in the author's doctoral dissertation on deposit at the University of Massachusetts library.

Durations and characteristics.—Of the interactions involving exposure durations and from one to all four stimulus characteristics, disregarding and with trial blocks, 6 of 13 first-order interactions, 6 of 10 second-order interactions, 5 of 8 third-order interactions, 3 of 4 fourth-order interactions, and 2 of 2 fifth-order interactions were significant. The highest order interaction was also significant. This interaction reflected a pattern of increasing increments from Trial Block 1 to 2 for increasingly easy combinations of both durations and characteristics. Most of the significant lower-order interactions involving durations and trials; characteristics and trials; and durations, characteristics, and trials can be regarded as special cases of this more general pattern. This pattern also appeared in most of the nonsignificant interactions, but less markedly.

DISCUSSION

The primary objective of the present experiment was to obtain parametric information about acquisition as functions of St and St-R durations. Findings pertinent to this objective are considered before joint effects of durations and stimulus characteristics.

Durations.—For St and St-R durations, disregarding stimulus characteristics, the family of curves for recall frequency and those for recall latency were mirror images. This relationship did not hold between recall-frequency and anticipation-frequency curves primarily because *S*s could not produce overt anticipations within the ½-sec. St duration combinations. The shortest response latency recorded during recall tests was 800 msec. Whether response latencies could be further reduced with practice so as to occur within the 500-msec. limit is conjectural.

Failure to anticipate did not mean that no learning had occurred as evidenced in Fig. 1 by the contrast between recall and anticipation curves for the ½-sec. St duration combinations as St-R durations lengthened. Except for combinations involving ½-sec. St durations, correlations were sufficiently high between response measures to suggest that only recall responses need be recorded. Furthermore, for short St durations, recall trials provide information on the progress of learning which cannot be obtained from anticipation trials.

The present finding that acquisition of PA units is a direct function of St and St-R durations is consistent with results of previous experiments. In the present experiment, disregarding stimulus characteristics, an eightfold increase in exposure durations from ½:½ to 4:4 sec. produced slightly less than an eightfold increase in total correct responses. Unfortunately no information is available on total time to learn in the present experiment.

In the present experiment, effects of Sim and M of stimulus and response members were less important per se than as providing a range of differences among combinations of stimulus characteristics in a rank order reasonably consistent with other findings for those same combinations. That this was accomplished, is evidenced by the rho of .68 between the present mixed list and four unmixed lists of Goss et al. (1962) which represented identical combinations of stimulus characteristics using different CVC trigrams.

Stimulus integration and response integration are among the processes which presumably underlie effects of Sim and M of stimulus and response members on PA learning (Goss, 1963; Goss & Nodine, 1963). Integration refers to the formation of stable recognition responses to stimuli in the form of distinctive spelling or pronunciation responses.

Once integration is complete, these recognition responses are evoked to stimulus and response members with a minimum of variability in form, latency, and duration increasing their availability for S-R association. In the present analysis the direct effects of M and inverse effects of Sim of stimulus and response members are regarded as reflecting differences in degree of integration of recognition responses to these stimuli.

Integration should also vary directly with frequency and duration of exposure of each stimulus and response member. Lengthening St durations should give *S*s more time per trial to rehearse each stimulus member hastening integration and, as a result, giving *S*s more time to anticipate.

Lengthening St-R durations should not only speed response integration by giving *S*s more time per trial to rehearse response members (and also stimulus members), but also give *S*s more time per trials to rehearse S-R connections.

Confirming these notions, of *S*s who experienced St-R durations of $\frac{1}{2}$, 1, 2, and 4 sec., 45%, 55%, 62%, and 82%, respectively, reported rehearsing during this period. Approximately 90% of these *S*s reported rehearsing response members rather than stimulus members or both, suggesting response integration as the primary basis of the direct relationship between acquisition and St-R duration. Few *S*s reported rehearsing during St durations which is consistent with the notion that in PA learning, all that is required is differential recognition responses to stimulus members. Such differentiation may oftentimes be merely noticing and responding to the particular one of the three letters of a member which is different from the letters of most or all of the other stimulus members.

Interactions of durations and stimulus characteristics revealed a pattern of increasing differences among combinations of Sim and M of stimulus and response members as combinations of St, St-R durations increased. This finding is contrary to an analysis based on stimulus and response integration which would predict decreasing differences among combinations of stimulus characteristics with increasing St, St-R durations due to increased opportunity for *S*s to form stable recognition responses to stimulus and response members.

The most plausible explanation for this contrary finding is that *S*s' mastery of the list under shorter St, St-R durations had not progressed sufficiently to reveal maximum differences among means of correct responses in 21 acquisition and 4 recall trials. Other aspects of the findings, such as *S*s' reports of rehearsal during St-R durations, are consistent with an integration interpretation.

REFERENCES

Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

Bugelski, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 409–412.

Goss, A. E. Comments on Professor Noble's paper. In C. N. Cofer & B. S. Musgrave (Eds), *Verbal behavior and verbal learning.* New York: McGraw-Hill, 1963.

Goss, A. E., Morgan, C. H., & Golin, S. J. Paired-associates learning as a function of percentage occurrence of response members (reinforcement). *J. exp. Psychol.*, 1959, **57**, 96–104.

Goss, A. E., & Nodine, C. F. *Meaningfulness, familiarization, and similarity in paired-associates learning.* New York: Academic Press, 1963.

Goss, A. E., Nodine, C. F., Gregory, B. N., Taub, H. A., & Kennedy, K. E. Stimulus characteristics and percentage of occurrence of response members as factors in paired-associates learning. *Psychol. Monogr.*, 1962, **76**(13, Whole No. 531).

Hovland, C. I. Experimental studies in rote-learning theory: VII. Distributed practice of paired-associates with varying rates of presentation. *J. exp. Psychol.*, 1949, **39**, 714–718.

Wilcoxon, H. C., Wilson, W. R., & Wise, D. A. Paired-associate learning as a function of percentage of occurrence of response members and other factors. *J. exp. Psychol.*, 1961, **61**, 283–289.

(Received August 29, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 107–108

SUPPLEMENTARY REPORTS

A FURTHER TEST OF A GENERAL UTILITY THEORY MODEL FOR PROBABILITY LEARNING

KENNETH B. LITTLE
University of Denver
YVONNE BRACKBILL
University of Colorado Medical School
ROBERT B. ISAACS AND NORMAN SMELKINSON
Johns Hopkins University

The predictive capacity of a general utility model for probability learning was tested in a 3-choice, noncontingent task for the range of major π values from .40 to .90. 12-yr.-old children were used as Ss. Accurate predictions were obtained for π values of .60 and .70. Asymptotic response levels at lower values of π were overestimated, while higher values were underestimated. Probability matching was obtained only at the higher levels of π. It was concluded that the utility of varying responses (or of correct responding) was not independent of the relative frequency of occurrence of the stimulus events as assumed by this and other utility theory models.

In a recent study (Little, Brackbill, & Kassel, 1962) the authors proposed a general utility model for probability learning in the multichoice, noncontingent case. A test of the model indicated that it predicted asymptotic response proportions adequately as the number of choices, k, was varied, but that its predictive capacity was questionable when the relative frequency of the most frequently appearing event, π_A, was varied. Since only two levels of π_A, .50 and .60, were used, a supplementary investigation was made to determine the range of π values through which the model could be expected to hold.

Method.—A detailed description of the procedure and experimental material is given in the study cited above. The *S*s for the present experiment were 60 12-yr.-old children. They were randomly assigned to six experimental groups of 10 *S*s each. The task consisted in *S*'s predicting which common English letter—A, E, or T—would appear on each card of a deck of 400 cards shown one at a time. Objective frequencies of occurrence of the stimulus events were as follows for the six groups: (Group 1) 40:35:25, (Group 2) 50:30:20, (Group 3) 60:25:15, (Group 4) 70:20:10, (Group 5) 80:15:05, and (Group 6) 90:08:02.

Results.—To predict response levels for any value of π under this utility model, it is necessary to obtain an estimate of α, the ratio of the marginal utility of correct responses to the marginal utility of varying responses. The formulas for estimating α and for predicting stable state response levels are given in Little et al. (1962). Alpha values were computed originally for each *S*, based on his performance on Trials 301–400. However, an analysis of variance showed that the differences among mean α values for the six groups of *S*s were not significant, $F\ (5, 53) = 2.53$. Therefore, in predicting final response level for any one group, the mean α for the remaining 50 *S*s was used. The use of the combined sample in this fashion permitted a more stable estimate of the population α. By using the 95% confidence limits of the average α thus obtained, it was also possible to specify the 95% confidence interval of each predicted response proportion.

Table 1 gives the α value for each group, the predicted asymptotic response proportion, p_p, the 95% confidence limits for p_p, and the observed response proportion, p_o, based on Trials 301–400. Since there were no signifi-

TABLE 1

PREDICTED AND OBTAINED ASYMPTOTIC RESPONSE LEVELS FOR SIX EXPERIMENTAL GROUPS

π_A	α	p_p	$p_{p\pm 2\sigma}$	p_o
.40	1.690	.389	.381–.397	.360
.50	1.636	.470	.446–.494	.442
.60	1.562	.542	.506–.578	.559
.70	1.600	.627	.577–.677	.609
.80	1.505	.684	.618–.750	.802
.90	1.513	.762	.692–.832	.870

cant differences between the means for Trials 301–350 and 351–400 for any group, stable response levels were assumed to exist.

Examining the data in Table 1, it is clear that only the response levels of the .60 and .70 groups were predicted with reasonable accuracy. All other response levels fall outside the 95% confidence intervals of the predicted proportions. At π_A values below .60, response levels are overestimated, and at values above .70, they are underestimated. A corrolary observation is that probability matching is approximated only with higher values of π_A. For all groups, the least frequently appearing event was chosen at a much higher rate than its actual frequency of occurrence. This accords with the findings of Brackbill and Bravos (1962).

From the results of this experiment it is clear that the assumption that α is independent of the objective frequencies is incorrect. The utility for *S* of varying his responses (or perhaps the utility of making a correct response) will shift with the probability of occurrence of the stimulus events. Examination of the average values of α for the six groups reveals a trend toward positive relationship with value of π, even though the differences among α means were not significant. Values of α for the six groups, beginning with the condition in which $\pi = .40$, were 1.06, 1.32, 1.70, 1.50, 1.98, and 1.94. If this trend is replicable, it should be possible to introduce a systematic correction into the model.

REFERENCES

BRACKBILL, Y., & BRAVOS, A. Supplementary report: The utility of correctly predicting infrequent events. *J. exp. Psychol.*, 1962, **64**, 648–649.

LITTLE, K. B., BRACKBILL, Y., & KASSEL, S. H. A test of a general utility theory model for probability learning. *J. exp. Psychol.*, 1962, **63**, 404–408.

(Received August 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 1, 108–110

RETENTION OF LEARNING IN A DIFFICULT TRACKING TASK

M. HAMMERTON[1]

Medical Research Council, Applied Psychology Research Unit, Cambridge, England

An experiment was carried out to examine the retention of skill in an acquisition task using 2nd-order (acceleration) control. 2 groups of Ss—each of 9 of the author's colleagues—were trained to a chosen criterion; 1 group then overlearned to a higher criterion. After 6 mo., both groups were retested, and brought up to the level of the 1st criterion. Both groups showed marked forgetting ($p < .04$) but the overlearning group still recalled better ($p < .05$). Both groups recovered very quickly. Conclusion is: if task is sufficiently difficult, overlearning does not prevent, but only somewhat reduces, decrement of performance with time.

In an extensive survey, Bilodeau and Bilodeau (1961) concluded that retention of a motor skill was virtually perfect for *Ss* who received prolonged practice. Retention remained perfect for periods of up to 2 yr. However, most studies with tracking tasks had used velocity or positional controls, and a very high degree of overlearning (e.g., Parker & Fleishman, 1962). Both velocity and positional controls are easy to handle; the present study deals with second-order (acceleration) controls which are much more difficult. In many practical training situations extensive overlearning is not possible, and a moderate degree was used in the present experiment. The task used involved aligning a moving spot of light on a cathode-ray tube on a stationary target spot.

Procedure.—Eighteen of the author's colleagues at the Applied Psychology Research Unit—16 men and 2 women—volunteered to act as *Ss*. Six of these had some experience of tracking. The *Ss* were divided into two groups of 9 each, which were balanced, as far as possible, for experience, age, and sex. The first day's performances by the two groups were later compared, using the *t* test. There was found to be no reliable difference. The two groups were designated A and B; it was

---

[1] The author wishes to thank his colleagues for their cooperation as *Ss*. The work was carried out under the general direction of D. E. Broadbent and E. C. Poulton.

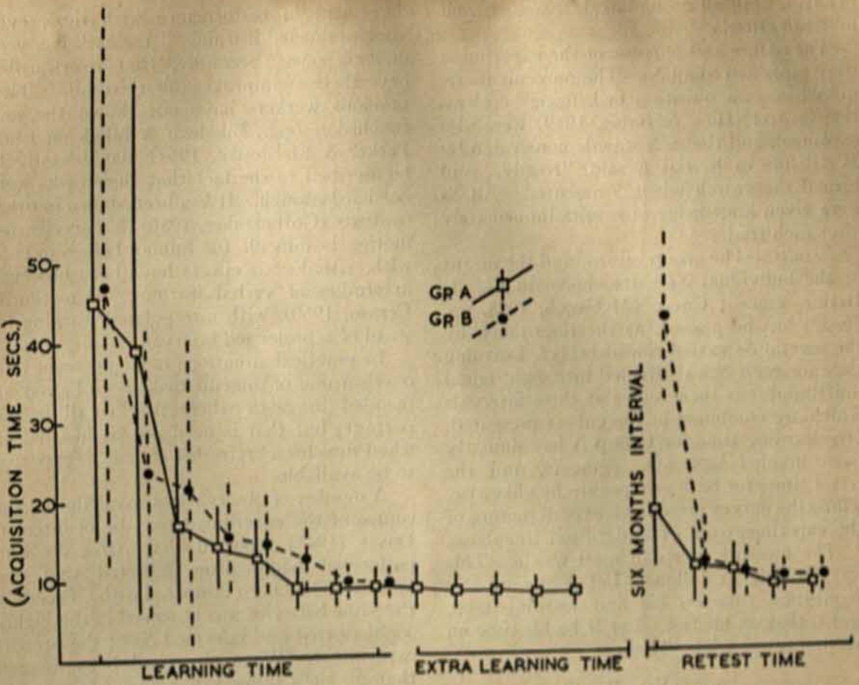


FIG. 1. Overall results of Groups A and B. (Variation of group means and means of *Ss'* *SDs* on standardized time-base.)

decided by the toss of a coin that Group A should overlearn.

The apparatus used was the control simulator which has been in use at the Applied Psychology Research Unit for some time. The *S* sat facing a cathode-ray tube upon which a short vertical line and a spot of light can be made to appear. The line is the target, and the movements of the spot are controlled by *S*, using a thumb joystick. An acceleration control was used; i.e., the acceleration of the spot across the screen was proportional to the deflection of the stick. The initial distance of the target from the spot was 22.5 mm. Each trial began when *E* turned a switch, which simultaneously moved the target to its position and started a dekatron timer. The spot had to be moved to the target, and kept within a zone 1.5 mm. on either side of it for 2 consecutive sec. When this had been done, the dekatron stopped, the joystick was made inoperative, and target and spot together automatically returned to the zero position. The reading of the timer, minus the 2-sec. "holding time," was *S*'s score. The display/control ratio of the system (rad sec^{-2}/rad) was 1.33 sec.^{-2}; the maximum possible linear acceleration of the spot across the screen was 8.9 cm. sec.^{-2}. The equipment was calibrated daily to insure constant conditions. Spot and stick movements were recorded with a pen recorder.

The attempt to make the task a difficult one was successful; one *S* described it as a "nightmare."

All *Ss* were given 5 trials per day until each had reached two criteria. These were: (*a*) Three successive daily mean scores below 12 sec.; (*b*) Three successive days' scores which did not differ significantly at the 5% level. This learning took various *Ss* from 8 to 22 days; i.e., from 40 to 110 trials. Trials with Group B *Ss* then stopped. Group A *Ss* were then given 10 trials per day until they had achieved 3 successive days' scores which did not differ significantly at the 1% level. This degree of overlearning required from 9 to 17 additional days; i.e., 90 to 170 additional trials. Whenever possible testing was on consecutive week days, although some *Ss* missed a few days.

For all *Ss*, retesting began 26 wk. after their last test. It was continued, at five trials per day, until all *S*s had again reached their first two criteria.

The nature and purpose of the experiment were explained to all *S*s. The movements required in an acquisition task using acceleration control (Hick & Bates, 1949) were also explained, and the task was demonstrated by *E*. Before each trial *E* said, "Ready," and turned the switch when *S* assented. All *S*s were given knowledge of results immediately after each trial.

Results.—The group means, and the means of the individual *SD*s are shown in Fig. 1. Here a Vincent Curve (McGeoch, 1948, pp. 36–38) has been used, as the time taken by the several *S*s varied considerably. Learning time for each *S* was divided into eight equal units, and it is their scores at these intervals which are combined in the curves presented. Overlearning time for Group A has similarly been divided into five segments, and the retest times for both groups into five likewise. Thus, the curves present an overall picture of the experiment on a standardized time-base.

The measure of recall used was ρ. This was defined as follows: Let F be an *S*'s acquisition time on his first learning trial, and L that on his last. Let R be his time on his first retest trial. Then:

$$\rho = \frac{\mathrm{F} - \mathrm{R}}{\mathrm{F} - \mathrm{L}}$$

ρ values varied from 1.00 to −2.15. Only one *S* obtained 1.00; it may be noteworthy that he was a qualified pilot. All *S*s together were ranked for ρ, and the Mann-Whitney *U* test (Siegel, 1956) was used to see if the two groups differed. It was found that $U = 18$, which is significant ($p < .05$; one-tailed). Thus overlearning improved initial recall. An examination of the curves, however, shows that the recall was far from perfect. Use of the sign test (Siegel, 1956) shows that the decrement is significant both for Group A ($p < .04$; two-tailed) and for Group B ($p < .004$; two-tailed).

For both groups recovery was very rapid; only three *S*s took more than 5 days to recover their previous performance level.

Discussion.—The present study seems to lead to two conclusions. First, that there is a decrement in performance with time, even after prolonged learning, if the task is a very difficult one. Secondly, that overlearning nevertheless improves first recall. That previous workers have not shown the first conclusion (e.g., Bilodeau & Bilodeau, 1961; Parker & Fleishman, 1962) may tentatively be ascribed to the fact that their tasks were not hard enough. It has been shown in other contexts (Gottsdanker, 1956) that accelerated motion is difficult for human beings to cope with. Analogous effects have been observed in studies of verbal learning (Slamecka & Ceraso, 1960), with interpolated learning instead of a prolonged interval.

In practical situations it would seem that overlearning of difficult tasks is to be recommended in cases where first recall is important; but that it need not be insisted on when time for a "refresher course" is expected to be available.

A number of observations made during the course of the experiment may be of interest. Davis (1948) reported that naive *S*s who performed badly often asserted that the apparatus had been tampered with. Precisely the same behavior was observed in the highly sophisticated and informed *S*s of the present experiment. When they had been assured that no such thing had happened, some *S*s supposed that they were being deceived for experimental reasons, and would assert without doubt that some parameter had been altered.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD. Motor skills and learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–279.

DAVIS, R. D. *Pilot error: Some laboratory experiments.* London: His Majesty's Stationary Office, 1948.

GOTTSDANKER, R. M. The ability of human operators to detect acceleration of target motion. *Psychol. Bull.*, 1956, **53**, 477–487.

HICK, W. E., & BATES, J. A. V. The human operator of control mechanisms. *Ministr. Supply Monogr.*, 1949, No. 17-204.

MCGEOGH, J. A. *The psychology of human learning.* London: Longmans Green, 1948.

PARKER, J. F., & FLEISHMAN, E. A. Factors in the retention and relearning of perceptual-motor skill. *J. exp. Psychol.*, 1962, **64**, 215–226.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

SLAMECKA, N. J., & CERASO, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–475.

(Received September 5, 1962)

DIGITAL LOGIC PACKAGES

BRS Electronics has available three series of digital logic packages which are designed to easily allow the assembly of logic networks. These networks can be designed to achieve precise control of various experimental parameters. The use of digital techniques and solid-state electronic circuitry gives precise timing, counting, programming, decision making, etc. The high-speed capabilities inherent to these units allows highly accurate time measurements in the microsecond range. Onset and termination of stimuli can be programmed to an accuracy in orders of magnitude greater than those available through electro-mechanical devices. The programming of the traditional, as well as unique, schedules of reinforcement is easily achieved. Exact control of electro-mechanical schedule-subject interface devices is obtainable.

The units marketed by BRS are known as DigiBits. They are an off-shoot of the experience gained by TECH SERV, Inc., the parent company, in the computer field and have been designed specifically for use by those individuals who do not have an extensive electronics background. The operation of DigiBits is easily grasped and the techniques required are acquired with little effort. Our technical staff will perform, at no charge, the necessary logic design work to translate your experimental requirements into a digital system. Send your requirements to the attention of Frank Goodwin

BRS *electronics*™
BEHAVIOR RESEARCH SYSTEMS
A TECH SERV CO.

5451 HOLLAND DRIVE
BELTSVILLE, MARYLAND
301-474 2900

Bi-monthly

$10.00 a year
($10.50 foreign)
$2.00 a copy

JOURNAL OF COMPARATIVE & PHYSIOLOGICAL PSYCHOLOGY

Editor
W. K. Estes
Stanford University
Associate Editor
Eliot Stellar
University of Pennsylvania
Consulting Editors
Gordon H. Bower
Stanford University
Joseph V. Brady
Walter Reed Army Institute of Research
Byron A. Campbell
Princeton University
Vincent G. Dethier
University of Pennsylvania
Frank A. Geldard
Princeton University
John F. Hall
Pennsylvania State University
Margaret Kuenne Harlow
Wisconsin Regional Primate Research Center
John I. Lacey
Fels Research Institute
Douglas H. Lawrence
Stanford University
Daniel S. Lehrman
Rutgers University
Frank A. Logan
Yale University
Mortimer Mishkin
National Institute of Mental Health
Conrad G. Mueller
Columbia University
Carl Pfaffmann
Brown University
W. N. Schoenfeld
Columbia University
Larry Stein
Wyeth Laboratories
Philip Teitelbaum
University of Pennsylvania
Hans-Lukas Teuber
Massachusetts Institute of Technology
J. M. Warren
Pennsylvania State University
David Zeaman
University of Connecticut

CONTENTS

Weiskrantz, Lawrence, and Cowey, Alan. Striate Cortex Lesions and Visual Acuity of the Rhesus Monkey........ 225

Gross, Charles G. Locomotor Activity Following Lateral Frontal Lesions in Rhesus Monkeys.................... 232

Wilson, William A., Jr., Oscar, Marlene, and Gleitman, Henry. The Effect of Frontal Lesions in Monkeys upon Widely Spaced Delayed-Response Trials................. 237

Warren, J. M., Grant, R., Hara, K., and Leary, R. W. Impaired Learning by Monkeys with Unilateral Lesions in Association Cortex.................................. 241

Stamm, J. S., and Knight, Marcia. Learning of Visual Tasks by Monkeys with Epileptogenic Implants in Temporal Cortex....................................... 254

Thompson, Robert. Thalamic Structures Critical for Retention of an Avoidance Conditioned Response in Rats...... 261

(continued on inside back cover)

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Journal of
Experimental Psychology

Vol. 66, No. 2 August 1963

RELATIONAL ISOLATION AS A MEANS OF PRODUCING THE VON RESTORFF EFFECT IN PAIRED-ASSOCIATE LEARNING[1]

RAYMOND L. ERICKSON[2]

University of California, Los Angeles

1 S-R pair in each of 8 9-pair lists of syllables and numbers was isolated solely on the basis of the relationship between its members. Stimulus and response elements of the isolated pairs were each individually more massed than isolated among their respective counterparts. An obtained isolation effect ($p < .001$) was interpreted to indicate that S-R intralist associative interference explanations based solely upon reduced stimulus and/or response generalization are inadequate to account completely for the von Restorff effect. Additional isolation was produced for some Ss by printing critical elements in red. Color isolation of both S-R elements produced an effect greater than that of response isolation alone ($p < .05$) but not greater than that of stimulus isolation alone. Stimulus and response color isolation did not produce different effects. Ss were 176 undergraduates.

The existence of the von Restorff, or isolation, effect has been clearly demonstrated. H. von Restorff (1933), using both serial and paired-associate lists, found that when relatively isolated items of one type were embedded in lists containing other "massed" material, the recall of the isolated items was superior. Subsequent investigations (e.g., Kimble & Dufort, 1955; Nachmias, Gleitman, & McKenna, 1961; Newman & Saltz, 1958; Pillsbury & Raush, 1943; Saltz & Newman, 1959; Saul & Osgood, 1950; Siegel, 1943) have confirmed the existence of an isolation effect in both paired-associate and serial learning.

Although the isolation effect is clearly demonstrable at an empirical level, marked differences have evolved with regard to the theoretical explanation of the phenomenon. The gestaltists contend that the effect is to be understood in terms of the gestalt laws of perceptual organization.

---

[1] Based on a dissertation submitted to the Psychology Department of the University of California, Los Angeles, in partial fulfillment of the requirements for the PhD degree. The author wishes to express his indebtedness to Norman H. Anderson for his advice and encouragement during the course of this study. A report on these data was made at the 1962 Western Psychological Association meeting in San Francisco.

[2] Now at the University of New Hampshire.

The basic precept of the gestalt approach is that, in the case of the isolation phenomenon as elsewhere, perceptual processes persist in and determine mnemonic processes (Koffka, 1935, pp. 481–506; von Restorff, 1933). Green (1958), on the other hand, maintains that the isolation effect does not depend upon structural *isolation* of items within a list. He contends that the "isolation" effect results from increased attention given any element that introduces a change into the structure of a list, so that in serial lists the subsequent massed material is not relevant to the effect. "Attention," for Green, is a primitive notion, similar in a sense to the gestalt concept of figure-ground, and he does not explain why increased attention should produce more rapid learning. Finally, current S-R explanations of the effect are based largely upon Gibson's (1940, 1942) principles of stimulus generalization and differentiation. They attribute the isolation effect to reduced intralist interference, resulting from reduced stimulus and/or response generalization (e.g., Kimble & Dufort, 1955; Postman & Phillips, 1954; Saul & Osgood, 1950; Siegel, 1943). Newman and Saltz (1958) and Saltz and Newman (1959), in particular, extended Gibson's principles to include response generalization and differentiation in an attempt to explain several aspects of the isolation effect in serial learning, although they were not satisfied with the success of their attempt.

Any S-R explanation that attributes the isolation effect solely to a reduction in intralist interference resulting from reduced stimulus and/or response generalization will be referred to here as the S-R intralist associative interference explanation. To date, no investigation has demonstrated an isolation effect in multitrial learning under such a condition that the S-R intralist associative interference explanation is clearly inapplicable. The major purpose of the present investigation was to demonstrate such an isolation effect by constructing paired-associate lists in which the only basis for isolation would be the *relationship* between the stimulus and response of each critical pair.

Table 1 shows two examples of the eight sets of items used in the lists. The critical pair in List 1 (KSC-ZNH) is isolated because both the stimulus and response are consonant syllables, whereas each of the other eight pairs of items in the list consists of a consonant syllable and a number, the syllables functioning as stimuli for half the eight pairs and as responses for the other half. The critical pair in List 2 (CTG-472) is isolated because it is composed of a consonant syllable and a number, whereas all the other pairs consist of pairs of consonant syllables or of numbers. In such lists, both the stimulus and the response of each critical pair are individually more massed than isolated among their respective counterparts; hence neither member of the critical pair is subject to reduced generalization tendencies.

An additional purpose of the present study was to investigate the differential effects of individually isolating either the stimulus, the response, or both elements of each critical pair along the dimension of color.

METHOD

Materials and apparatus.—Eight basic nine-pair lists, each containing a different relationally isolated pair, were used. Each isolated pair appeared in one of the other lists as a nonisolated pair.

List 3 was identical to List 1 (in Table 1), except that a number-number pair constituted the isolated pair. List 4 was identical to List 2, except that a number-syllable combination functioned as the isolated pair. Lists 5 through 8 replicated the first four, except that different syllables and numbers were used. Hence, there was a complete counterbalancing of syllable-syllable, syllable-number, number-number, and number-syllable combinations in the isolated position.

Each relationally isolated pair of items appeared under four color conditions: (*a*) Black-Black: both stimulus and response in black; (*b*) Red-Black: stimulus in red, response in black; (*c*) Black-Red: stimulus in black, response in red; (*d*) Red-Red: both stimulus and response in red. All nonisolated pairs appeared in black only. Condition Black-Black thus provided the experimental condition for evaluating the relational isolation effect while the other three color conditions permitted evaluation of the differential stimulus-response isolation effects.

All consonant syllables used came from the 50% association level of Witmer's (1935) list of association values. The lists were printed on 80-lb. fotolith paper by means of rubber stamps with symbols ⅜ in. in height. Each list appeared on an endless tape in nine different orders, with no spaces separating the orders. A Gloric memory drum with a 2-2-1 sec. time sequence was used.

Design and procedure.—The first replication of the experiment used 128 *S*s assigned in equal numbers to each color condition. In the second replication, which was temporally contiguous with the first and had as its major purpose acquiring additional data for Cond. Black-Black, 32 additional *S*s were run under Cond. Black-Black and an additional 16 *S*s were run under Cond. Red-Red as a check on replication effects. Lists were counterbalanced within color conditions in both replications.

Each *S* first practiced to a criterion of one perfect recitation on a list of six pairs of two-syllable nouns. He was then presented with his appropriate experimental list, on which learning continued to a criterion of one errorless trial or 25 presentations, whichever took longer. The *S*s were instructed to spell out their responses to the consonant syllables, and were told not to be afraid to guess and make mistakes. Responses were recorded verbatim.

In addition to the above independent variables, three minor variables were included in the design. The *S*s were divided into four groups on the basis of their College Entrance Examination Board (CEEB) scores, and the *S*s from each CEEB level were randomly assigned to the 32 list-color conditions. Eight of the nine different orders of each list were used as initial orders of presentation, and these eight starting orders were counterbalanced within lists, color conditions, and CEEB levels. Finally, each *S* was specifically assigned to be run among the first, second, third, or fourth quarter of the *S*s and these temporal groups were counterbalanced in a like fashion. For a more detailed description of procedure, see Erickson (1962).

TABLE 1

TWO OF THE EIGHT SETS OF ITEMS USED IN THE EXPERIMENTAL LISTS

List 1	List 2
SWJ-217	SWJ-DXR
BJN-821	BJN-GDP
RKD-764	RKD-JPZ
CTG-472	KSC-ZNH
KSC-ZNH[a]	CTG-472[a]
086-DXR	086-217
590-GDP	590-821
305-JPZ	305-764
953-LFS	138-649

Note.—Lists were not presented in the order shown here. An original random order of the pairs was determined and nine subsequent arrangements of this random order were selected such that each pair appeared once at each of the nine positions and there was only one repetition of any adjacent pairs in running through 81 presentations. The nine orders were the same for all lists.

[a] Isolated pairs. The basis for isolation is the S-R *relationship* between the items of the critical pair. In List 1, for example, KSC-ZNH is a syllable-syllable pair in a list otherwise composed of syllable-number and number-syllable pairs.

The design confounded certain higher order interactions of these three minor variables. Since these interactions would be expected to be negligible on a priori grounds, and since the main effects were not significant (with the single exception of a barely significant *F* for CEEB level), the procedure seems justified.

Subjects.—The *S*s were 176 volunteer undergraduate psychology students at Whittier College. There were 56 males and 120 females.

RESULTS

The dependent variables extracted for analysis were: (*a*) number correct responses to the isolated pair over Trials 1–25; (*b*) mean number correct responses to the nonisolated pairs over Trials 1–25. The "nonisolated pairs" were defined for each list as those pairs whose stimulus and response elements were used somewhere

TABLE 2

Mean Correct Responses per *S* per Pair over Trials 1–25 for Isolated and Nonisolated Pairs in Each of the Four Color Conditions

Color Cond.	Isolated Pair	Nonisolated Pairs
Black-Black	10.23	6.92
Red-Black	16.28	7.58
Black-Red	14.06	6.98
Red-Red	16.96	6.92

in other lists as elements in isolated pairs. Consequently, six rather than eight nonisolated pairs were scored for each list, the remaining two pairs being ignored.

Table 2 shows mean correct responses for isolated and nonisolated pairs in each of the four color conditions. Figure 1 presents the learning curves for isolated and nonisolated pairs within each of the four color conditions.

The most important results are from the Black-Black condition. As both Table 2 and Fig. 1 indicate, the isolated pairs were learned more easily than the nonisolated pairs under Cond. Black-Black. The difference of 3.31 correct responses is significant, as shown by the *F* ratio for Isolation in the analysis of variance of Table 3.

Variations along the color dimension produced no detectable effect on the ease of learning the nonisolated pairs, but did produce effects on the ease of learning the isolated pairs, and hence on the magnitude of the isolation effect. An analysis of variance of isolated scores alone yielded a significant *F* for color effects, $F(3, 144) = 12.23, p < .001$, indicating that at least one of the four means for isolated pairs shown in Table 2 differed significantly from the others. A similar analysis of nonisolated scores alone indicated no significant differences among the four nonisolated score means shown in Table 2, $F(3, 144) = 0.23$.

A more precise indication of color

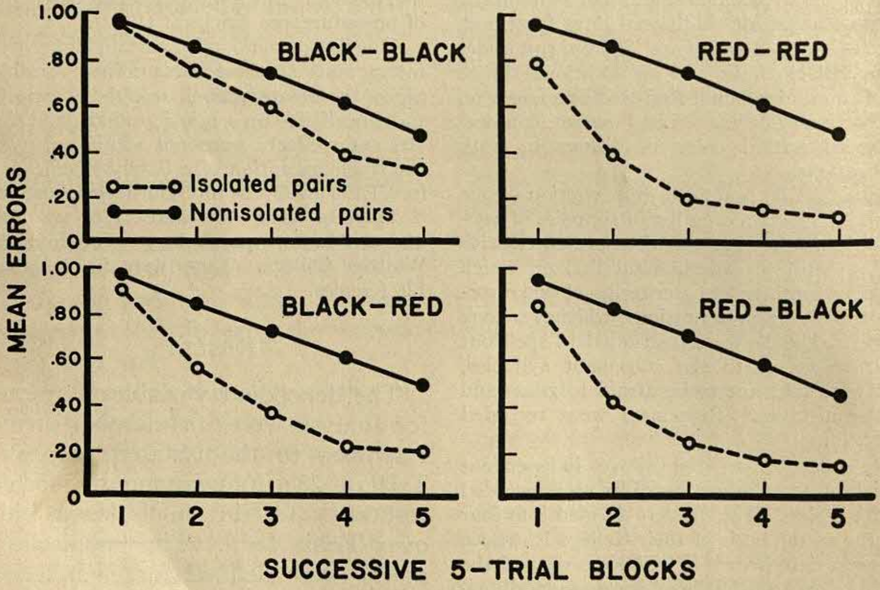


Fig. 1. Mean errors per *S* per pair per trial made to isolated and nonisolated pairs averaged over five-trial blocks for each color condition.

effects was obtained by applying a Duncan multiple-range test to the four color condition group means of the isolated scores. The mean isolated score under Cond. Black-Black was significantly lower ($p < .05$) than each of the other three means, but the means of conditions in which color was used did not differ significantly from each other.

Since the analysis of variance of nonisolated scores alone had shown that the use of color produced no detectable effect on the rate of learning the nonisolated pairs, it was considered appropriate to use the nonisolated scores as a covariable for the three conditions in which color was used. A second Duncan multiple-range test analyzed the differences among these three means of the isolated scores as adjusted by the covariance analysis. The adjusted mean of Cond. Red-Red was significantly larger ($p < .05$) than that of Cond. Black-Red, but no other differences were significant.

TABLE 3

ANALYSIS OF VARIANCE OF TOTAL CORRECT RESPONSES OVER TRIALS 1–25 FOR ISOLATED AND NONISOLATED PAIRS UNDER COND. BLACK-BLACK

Source	*df*	*F*
Between *S*s		
Lists (L)	7	1.08
Replications (R)	1	.03
L × R	7	1.16
Error (b)	48	(1,474.71)
Within *S*s		
Isolation (I)	1	21.48***
I × L	7	.93
I × R	1	.54
I × L × R	7	1.59
Error (w)	48	(588.59)

Note.—Error *MS*s in parentheses. For computational convenience, isolated scores were multiplied by 6 rather than dividing the nonisolated scores by 6.
*** $p < .001$.

In summary, any addition of color to the critical pairs of items produced a significant increase in the magnitude of the isolation effect compared to that obtained from relational isolation alone. The isolation effect under Cond. Red-Red was found to be greater than that under Cond. Black-Red, but not greater than that obtained under Cond. Red-Black. Conditions Red-Black and Black-Red did not produce detectably different effects.

As Table 3 shows, the Lists × Isolation interaction for Cond. Black-

TABLE 4

ANALYSIS OF MEAN ERRORS PER *S* OVER TRIALS 1–25 FOR ISOLATED PAIRS IN EACH OF THE FOUR COLOR CONDITIONS

Color Cond.	Incorrect Responses to Critical Stimulus							Incorrect Elicitations of Critical Response	
	Total Errors	No Response	Extralist Responses	Partial Correct Responses[a]	Partial Incorrect Responses[b]	Homogeneous Intrusions[c]	Heterogeneous Intrusions[c]	By Homogeneous Stimuli[d]	By Heterogeneous Stimuli[d]
Black-Black	14.77	9.82	.50	2.34	.96	.31	.84	.78	.25
Red-Black	8.72	5.06	.22	2.19	.75	.34	.16	.09	.44
Black-Red	10.94	7.97	.19	1.75	.69	.03	.31	.69	.34
Red-Red	8.04	3.55	.40	3.21	.40	.42	.06	.08	.29

[a] Partial correct responses defined as responses containing two elements from the critical response.
[b] Partial incorrect responses defined as responses containing two elements from any noncritical response.
[c] Homogeneous intrusions defined as intralist responses from same class as critical response; heterogeneous from opposite class. (Classes were numbers and syllables.)
[d] Homogeneous stimuli defined as stimuli from same class as critical stimulus; heterogeneous from opposite class.

Black was not significant. Hence, the highly significant *F* obtained for relational isolation cannot be attributed to idiosyncratic characteristics of particular lists. The only significant List effect in any of the analyses was a Lists × Isolation interaction, $F(7, 144) = 2.78$, $p < .01$, obtained in an analysis of scores from all four color conditions. Inspection of isolated and nonisolated means for each list showed that the isolation effect was in the same direction for all lists and that the significant interaction reflected the fact that the isolation effect was greater for some lists than for others. The cause of this difference is not known.

Table 4 presents an analysis of the errors made to isolated pairs. Of interest is the high proportion of "no response" to the critical stimulus; intralist intrusions constituted only about 6% of the total errors. Although the isolated pairs were learned more rapidly under Cond. Red-Red than under Cond. Black-Red, there were fewer intralist intrusions to the critical stimulus under Cond. Black-Red then under Cond. Red-Red, whether these are measured in absolute numbers or as proportions of total errors.

It is also of interest to note that Cond. Black-Red did not produce fewer elicitations of the critical response by inappropriate stimuli than did Cond. Black-Black. This holds true even when only incorrect elicitations by heterogeneous stimuli (i.e., stimuli associated with responses from which a maximum of response generalization could be expected) are considered. On the other hand, trends evident in the intralist intrusion data are consistent with the interpretation that both stimulus and response color isolation produced a greater reduction in intralist interference resulting from like elements than that resulting from unlike elements.

The effects of the minor variables require only brief comment. Analyses of variance of the main effects of each minor variable were carried out separately on isolated and nonisolated scores. All effects were found to be nonsignificant with the single exception of CEEB level on nonisolated scores, which was just significant at the .05 level. However, there was no orderly relationship among the performances at the four CEEB levels. No significant *F*s were obtained for replications, nor for any replication interactions. Hence, combining data from the two replications was considered to be appropriate.

DISCUSSION

The results of the present investigation are best considered in two parts: those from Cond. Black-Black, and those from the other three color conditions. The Black-Black results will be considered first.

The important finding here was that relational isolation facilitated the learning of the critical pairs. Pairs of items that possessed a unique S-R *relationship* within a list were learned more readily than pairs of items not possessing this unique relationship. Both gestalt theory and Green's attention hypothesis can easily account for this effect. According to the gestalt explanation, the trace systems of the relationally isolated pairs would constitute a figure on the ground of the trace systems of the nonisolated pairs. The isolated pairs would be learned more easily because their unique relationship causes their traces to "stand out" as figure, while the traces of the nonisolated pairs become assimilated with each other into a ground. If it is assumed that the relationally isolated pairs constitute a perceptible "structural change" in the lists, according to Green they would then receive increased attention and would be learned more rapidly for that reason.

However, the traditional S-R intralist associative interference explanation does not account for the relational isolation effect. According to this explanation, the more differentiated a stimulus item is from other list stimuli, the fewer inappropriate responses that stimulus will elicit. This follows because its greater differentiation (isolation) is equivalent to less stimulus generalization. Similarly, a response item that is more differentiated from the remainder of the response items than the latter are from each other will evoke less response generalization. Consequently, the stimulus for the isolated response item will elicit fewer responses that are appropriate elsewhere in the list. The isolation effect in paired-associate learning is held to result, then, from reduced intralist interference, the basis for which is either reduced stimulus generalization or reduced response generalization or both.

It is apparent that such an S-R explanation cannot account for the relational isolation effect found here. Under Cond. Black-Black, both the stimulus and the response items of the isolated pairs were individually more massed than isolated among their respective counterparts. There is thus no basis for assuming a reduction in either stimulus or response generalization for the critical pairs. It seems clear, therefore, that even if the more rapid learning of the relationally isolated pairs is attributed to a reduction in intralist interference, explanatory principles in addition to stimulus and/or response generalization will have to be utilized to account for this reduction in interference. It would be premature to infer that the results cannot be explained within the framework of general S-R theory, but account must somehow be taken of the pattern of relationships obtaining between the stimuli and the responses making up a given list.

One additional explanatory S-R principle would involve the concepts of mediating response and response-produced cues. Such a mediating response hypothesis would hold that the relational isolation effect is produced by a covert response made to the unique characteristic of the relationally isolated pair of items. The precise nature of this mediating response would be impossible to specify at this point. However, it would be assumed that the mediating response, whether conscious and well-articulated or not, is a response made to the relationally isolated pair of items because the presentation of that critical pair disrupts the tendency to respond to that type of stimulus item with the *type* of response that has received reinforcement throughout the rest of the list. It is assumed that this response functions as a typical mediating response, in that it provides discriminable cues. While the original evocations of the mediating response depend upon the prior appearance of the response item of the relationally isolated pair, the mediating response comes, as learning progresses, to be associated with the critical list stimulus alone. One chain of associations, then, is from the critical stimulus item to the mediating response, and from the response-produced cues to the critical response item.

The distinctive characteristic of the hypothesized mediating response is that it is a response brought about by a particular attribute of the list. Mediating responses utilized by *S*s in learning the nonisolated pairs, on the other hand, would depend upon each *S*'s idiosyncratic associations to particular elements and pairs. The advantage for the isolated pair of items results from the fact that the list structure makes it possible for all *S*s to share a common mediating response to the critical pair.

The main result from the remaining three color conditions was that any addition of color increased the size of the isolation effect. This raises some interesting questions. The most parsimonious S-R explanation would be to attribute the increased effect to a reduction in stimulus and/or response generalization, since with any introduction of color, either the stimulus or the response element of the critical pair, or both, were individually isolated. However, the error data relevant to Cond. Black-Red and Red-Red presented in Table 4 seem inconsistent with an interpretation that ascribes increased ease of learning of the isolated pairs solely to reduced stimulus and/or response generalization. Furthermore, there is no logical reason to assume that the factors that produced an isolation effect under Cond. Black-Black ceased to operate in the three conditions in which color was used. Granted the appropriateness of the mediating response hypothesis to account for the effect obtained in Cond. Black-Black, it may be assumed that similar mediating responses were operating in the other three color conditions. According to this argument, the addition of color may not only have reduced stimulus and/or response generalization, but it may also have changed the characteristics of the mediating response in that *S*s were now responding to the atypical color as well as to the difference in relationship of the critical pair. This could either have made the mediating response more discriminable, or increased the probability of its occurrence, or both.

Such a use of the mediating response hypothesis as an additional explanatory principle is related to an interpretation made by Gleitman and Gillett (1957), who proposed that the von Restorff effect may be a composite of two factors. The first of these they held to result from *S*s' deliberate efforts to organize the materials. They maintained that while the cues for organization may be inherent in the materials, their utilization perhaps depends upon intentional structuring, in a manner akin to the use of rhythm and other organizational devices. The second factor they presumed to depend upon lowered interference effects from other items in the list, and to operate regardless of intention. While mediating responses of the type hypothesized here need not depend upon intentional structuring, when intent is involved they would appear to be closely akin to what Gleitman and Gillett call organizing responses.

The results of the present investigation, together with those of Newman and Saltz (1958) and Saltz and Newman (1959), indicate that the von Restorff effect is more complex than intralist interference theories based solely upon the principles of stimulus and/or response generalization would allow. Since the existence of mediating responses in verbal learning seems well established (e.g., Bugelski, 1962; Bugelski & Scharlock, 1952; Russell & Storms, 1955; Underwood & Schulz, 1960, pp. 296–300), the mediating response explanation for the relational isolation effect found here has a certain plausibility. The extent to which mediating responses may be involved in all isolation phenomena remains to be determined.

REFERENCES

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 409–412.

BUGELSKI, B. R., & SCHARLOCK, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.

ERICKSON, R. L. Further investigation of the von Restorff effect in paired-associate learning: Relational isolation. Unpublished doctoral dissertation, University of California, Los Angeles, 1962.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

GIBSON, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.

GLEITMAN, H., & GILLETT, E. The effect of intention upon learning. *J. gen. Psychol.*, 1957, **57**, 137–149.

GREEN, R. T. Surprise, isolation and structural change as factors affecting recall of a temporal series. *Brit. J. Psychol.*, 1958, **49**, 21–30.

KIMBLE, G. A., & DUFORT, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50**, 361–368.

KOFFKA, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.

NACHMIAS, J., GLEITMAN, H., & MCKENNA, V. The effect of isolation of stimuli and responses in paired associates. *Amer. J. Psychol.*, 1961, **74**, 452–456.

NEWMAN, S. E., & SALTZ, E. Isolation effects: Stimulus and response generalization as explanatory concepts. *J. exp. Psychol.*, 1958, **55**, 467–472.

PILLSBURY, W. B., & RAUSH, H. L. An extension of the Koehler-Restorff inhibition phenomenon. *Amer. J. Psychol.*, 1943, **56**, 293–298.

POSTMAN, L., & PHILLIPS, L. W. Studies in incidental learning: I. The effects of crowding and isolation. *J. exp. Psychol.*, 1954, **48**, 48–56.

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

SALTZ, E., & NEWMAN, S. E. The von Restorff isolation effect: Test of the intralist association assumption. *J. exp. Psychol.*, 1959, **58**, 445–451.

SAUL, E. V., & OSGOOD, C. E. Perceptual organization of materials as a factor in influencing ease of learning and degree of retention. *J. exp. Psychol.*, 1950, **40**, 372–379.

SIEGEL, P. S. Structure effects within a memory series. *J. exp. Psychol.*, 1943, **33**, 311–316.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

VON RESTORFF, H. Über die Wirkung von Bereichsbildungen im Spurenfeld. In W. Köhler & H. von Restorff, Analyse von Vorgängen im Spurenfeld. I. *Psychol. Forsch.*, 1933, **18**, 299–342.

WITMER, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–360.

(Received September 13, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 120–126

INVARIANCE OF SIGNAL DETECTABILITY OVER STAGES OF PRACTICE AND LEVELS OF MOTIVATION[1]

JOHN A. SWETS[2] AND SUSAN T. SEWALL

Massachusetts Institute of Technology

Effects of practice and of different levels of motivation were examined in 2 psychophysical tasks. Under ordinary motivation the effect of practice in a forced-choice task was limited to the 1st session; under substandard motivation a small practice effect (equivalent to approximately 1.5 db. of signal power) extended over 3 sessions. Manipulation of the level of motivation after the practice period produced a very small effect—less than 1.0 db. in a rating task, and less than .5 db. in the forced-choice task. The relationship of these results to previous experimental results, and to a psychophysical application of statistical learning theory, is discussed.

It is generally accepted by psychophysicists working with the methods and measures of modern detection theory that learning and motivation have a very small effect on signal detectability. Specifically, it is taken as a rule of thumb that under ordinary experimental conditions the effects of practice are limited to a few minutes during *O*'s first acquaintance with his task, and that normal fluctuations in the level of motivation have a negligible influence on detection performance.

This assumption of stability has been based largely on a gradual accumulation of unpublished, and often informal, observations in several laboratories. The experiment described here was undertaken to provide additional formal evidence to supplement the few studies reported in the literature.

The results are also of interest in connection with a recent attempt to extend statistical learning theory to account for certain results in psychophysics (Atkinson, 1963). In Atkinson's model, variation over the usual experimental range in the detectability index d', under both forced-choice and yes-no detection procedures, is coordinated with experimental manipulations of the general level of motivation. According to detection theory and practice, changes in d', in any single task, are determined almost exclusively by changes in signal energy (Swets, 1961; Swets, Tanner, & Birdsall, 1961; Tanner & Swets, 1954).

It should be mentioned, principally to point out that this paper does not deal further with the topic, that detection theorists have occasionally been charged, justly, with ignoring or minimizing the learning effects exemplified by the dependencies in sequences of detection responses (Atkinson, 1963; Speeth & Mathews, 1961). It is understandable, of course, that learning theorists, accustomed to distinguishing among different learning models on the basis of their differential predictions for sequences of responses, would be more sensitive to the existence of sequential effects in psychophysical data. However that may be, it is fair to say that most psychophysicists would grant that sequential effects are potentially capable of providing information about the detection process. On the other hand, these effects too are typically very small under usual laboratory conditions, and it is unlikely that they are of substantial importance in investigations of substantive problems in sensory psychology.

---

[1] The experiment reported was conducted in the Research Laboratory of Electronics, Massachusetts Institute of Technology, with support from the United States Air Force (Operational Applications Laboratory and Office of Scientific Research), Army Signal Corps, and Office of Naval Research. This paper was prepared at Bolt Beranek and Newman Inc. with support from the Office of Advanced Research and Technology of the National Aeronautics and Space Administration.

[2] Also at Bolt Beranek and Newman Inc., Cambridge, Massachusetts.

METHOD

The experiment was divided into four successive experimental conditions. In the first, when *O*s were naive, the major interest was in the effect of practice. In the remaining three conditions, motivation was the principal variable.

Throughout the experiment the signal to be detected was a .1-sec. burst of a tone of 1,000 cps. The signal was masked by a continuous white noise with a spectrum level of approximately 50 db. re .0002 dynes/cm^2. In order to identify the signal clearly for *S*s, before every block of 100 trials they were given 5 trials with the masking noise attenuated by 10 db.

The *S*s observed 2 hr. a day, with rest periods interspersed frequently, 5 days a week. Each 2-hr. session contained 800 trials. The *S*s were college students. They were paid $1.30 per hour.

Two psychophysical procedures were used: the two-alternative, forced-choice procedure (2AFC), and the rating procedure. On each trial of the 2AFC procedure two temporal intervals were marked by lights. The signal occurred in either the first or the second interval with equal probability. Following the two intervals, *O* pressed one of two buttons to indicate the interval in which he believed the signal occurred. He was given knowledge of results immediately by means of lights. The trial sequence was as follows: warning light, .1 sec.; space, .2 sec.; first signal interval, .1 sec.; space, .8 sec.; second signal interval, .1 sec.; space, .1 sec.; answer period, 1.5 sec.; feedback, .2 sec.; space, .4 sec. The total time of a trial was 3.5 sec. The dependent variable in this task is the proportion of correct responses, $P(c)$, which can be simply converted to the detectability index d' (see Swets et al., 1961).

On each trial in the rating procedure the signal occurred with a probability of .5 in a single temporal interval. The *O* responded by means of four buttons which were ranked to indicate his confidence that a signal was present. The categories of confidence were defined as a "certain yes," an "uncertain yes," an "uncertain no," and a "certain no." These responses generate three points on an operating-characteristic curve, and thus an estimate of d'. The details of the curve-fitting process will not be discussed here since they have been described in detail by Egan, Schulman, and Greenberg (1959). The trial sequence again occupied 3.5 sec. and was divided as follows: warning, .1 sec.; space, .2 sec.; signal interval, .1 sec.; space, 1.0 sec.; answer period, 1.5 sec.; feedback, .2 sec.; space, .4 sec.

In Cond. 1, three *O*s served for 5 successive days under the 2AFC procedure. On each day a block of 200 trials was presented at each of four signal energies. These signal energies —stated in terms of 10 log E/N_o, where E is the signal energy, or time integral of power, and where N_o is the noise power in a 1-cycle band—were 9, 10, 11, and 12; they yield values of $P(c)$ ranging from approximately .60 to .90. The order of presentation of the four signal levels was chosen randomly each day. In the remainder of the experiment, just one signal energy was used, namely, 10 log $E/N_o = 10$, which yields a $P(c)$ in the neighborhood of .75.

Prior to Cond. 2, 3 days of practice were given in the use of the rating procedure. Conditions 2, 3, and 4 each consisted of 3 days under the rating procedure (2,400 trials) followed by 1 day under the 2AFC procedure (800 trials). Conditions 2, 3, and 4 represent, respectively, three levels of motivation, termed ML I, ML II, ML III. Roughly, in terms of experimental practice, these three levels of motivation were substandard, standard, and suprastandard. The top of Table 1 summarizes the experimental design as just described.

Motivation Level I was achieved by indoctrinating *O*s in a relatively casual fashion. Specifically, they were hired, instructed, and put to work by the young lady who conducted the experiments. She gave them the impression that she was employed as a technical assistant to conduct experiments for a professor who was rarely in evidence around the laboratory. She conducted herself efficiently, but attempted to show no particular interest in the outcome of the experiment. She

TABLE 1

SUMMARY OF THE EXPERIMENTAL DESIGN

	Cond. 1	Cond. 2		Cond. 3		Cond. 4	
	Days 1–5	Days 9–11[a]	Day 12	Days 13–15	Day 16	Days 17–19	Day 20
Group A	2AFC	rating	2AFC	rating	2AFC	rating	2AFC
	ML I	ML I		ML II		ML III	
	4 signal energies	1 signal energy					
Group B	2AFC ML II 4 signal energies						

[a] Days 6–8 were devoted to practice under the rating procedure.

presented Cond. 1 and 2 as a training period.

Following Cond. 2 under ML I, ML II was effected in Cond. 3 by the appearance of the heretofore absent professor, who spent 30 min. discussing with *O*s the general purpose of his experiments. He pointed out that the results were of interest with respect to theory in psychology; that the laboratory was supported by the United States Army Signal Corps, the Air Force, and the Navy; that these organizations based their detection procedures on the results of such experiments; and he solicited their cooperation. He proceeded to discuss with them certain specific aspects of experimental procedure. He told them something about random sequences in an attempt to educate them about the gambler's fallacy. He asked them to delay making a choice under the 2AFC procedure until they had attended carefully to both intervals, and to try to hold fairly stable criteria for their confidence judgments under the rating procedure. He indicated a concern to deal with any questions they might have about procedure. Motivation Level II thus reflects the care and precautions ordinarily taken by an *E*.

Motivation Level III was instituted in Cond. 4 by promising *O*s a relatively large monetary reward for improved performance. They were told in general terms of the existence of a theory of ideal *O*s (Swets, 1961), and it was made explicit that the interest at that point was in just how well they could do under the best conditions. The bonus was presented to add to their incentive. The bonus under the 2AFC procedure was $1.00 for each increase in $P(c)$ of .01 relative to each *O*'s best performance up until that time. Thus, an increase in $P(c)$ from .75 to .85 on the day devoted to 2AFC would earn a bonus of $10.00. Bonuses under the rating procedure were given on the basis of increments in d'. The *O*s were informed that their performance was measured in terms of d' and that an increase in d' of .05 was equivalent to an increase of .01 in $P(c)$ under 2AFC. Accordingly, they would receive an extra $1.00 per day for each increment of .05 in d' on the 3 days devoted to the rating procedure. So an increase of .5 in the d' obtained from the ratings would bring a return of $10.00 a day. Thus, what might be considered an attainable improvement—namely .5 in d'—would raise *O*'s income for the 4 days of Cond. 4 from about $10.00, based on the hourly rate, to about $50.00.

To determine the interaction of practice and motivational effects, a second group of three naive *O*s was employed only in Cond. 1; that is, for 5 days they observed four signal energies under the 2AFC procedure. As indicated in Table 1 these *O*s, unlike those in the first group, practiced under the standard level of motivation, ML II.

RESULTS

Practice.—Figure 1 shows the average results for Cond. 1 obtained from the three *O*s under ML I. (We present the average result of the three *O*s for the sake of brevity; no individual differences are discernable.) The average values of $P(c)$ for the four signal energies are for Days 1 through 5: .69, .75, .78, .77, and .76, respectively. The corresponding values of d' are .71, .96, 1.07, 1.05, and 1.00. Given the empirical finding that d' is approximately proportional to signal power (Green, Birdsall, & Tanner, 1957), the improvement from Day 1 to Day 3 is equivalent to an increase of 1.8 db. in signal power. The improvement from Day 1 to Day 4 is less (1.65 db.), and, from Day 1 to Day 5, less still (1.5 db.).

TABLE 2

VALUES OF d' OBTAINED WITH THE RATING PROCEDURE UNDER THREE LEVELS OF MOTIVATION

O	Practice	ML I	ML II	ML III
1	.72	.75	.81	1.00
2	.63	.68	.71	.99
3	.74	.83	.93	1.08
Average	.70	.75	.82	1.02

Experience suggests that "experimental error" can account for no more variation than .5–.75 db. with the forced-choice procedure, so the improvement shown here is not entirely inconsequential. However, the second group of *S*s, trained under the standard level of motivation (ML II), did not show any improvement. Their results are displayed in Fig. 2. For both Days 1 and 5 the average $P(c) = .80$ ($d' = 1.20$). The average result ranges from $P(c) = .78$ to $P(c) = .80$, from $d' = 1.10$ to $d' = 1.20$—a range equivalent to .4 db. (Again, the average of the group does not mask individual differences. Comparing Days 1 and 5, the changes in $P(c)$ for the individual *O*s were $+.01$, $-.01$, and 0.0. In fact, the slight decrement in the average result on Days 2 and 4 relative to Days 1, 3, and 5 was shown by each *O*.)

Motivation.—The results obtained with the rating procedure in Cond. 2, 3, and 4 are shown in Table 2. The entries are values of d'. The average improvement from ML I to ML II is the equivalent of .4 db., and from ML II to ML III, the equivalent of .9 db. These changes in the values of

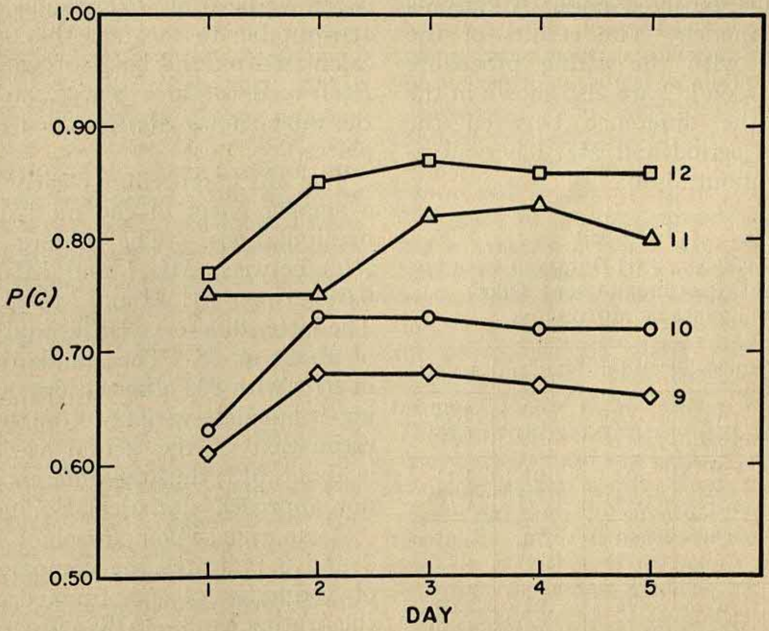


FIG. 1. Effect of practice under a low level of motivation.

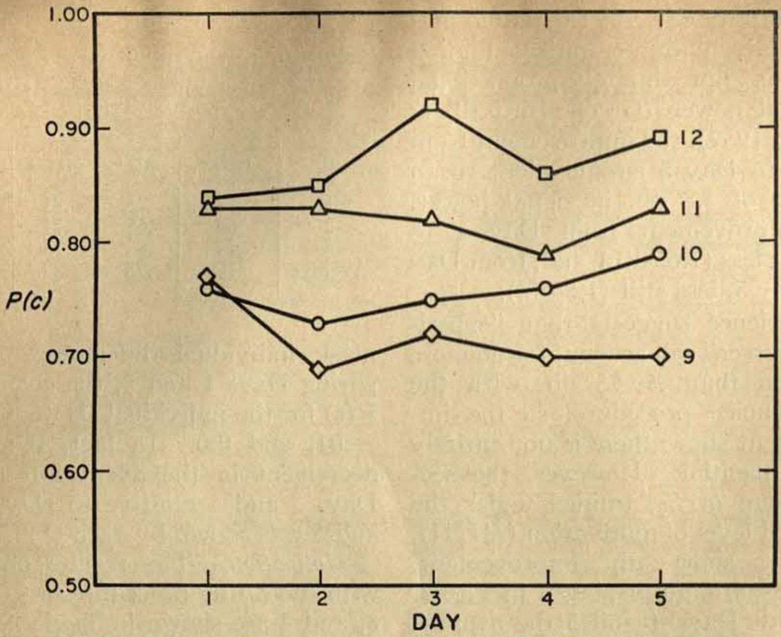


FIG. 2. Effect of practice under a standard level of motivation.

d' correspond to changes in $P(c)$ under 2AFC from .70 to .72, and from .72 to .76. The largest effect, from ML I to ML III, is .27 in d'—about 1/10 of the range expected under Atkinson's (1963) model. The results of the practice with the rating procedure prior to Cond. 2 are also shown in the table; the difference between the practice period and ML I is equivalent to about .3 db.

TABLE 3

VALUES OF d' AND $P(c)$ OBTAINED WITH THE 2AFC PROCEDURE UNDER THREE LEVELS OF MOTIVATION

O	Index	ML I	ML II	ML III
1	d' $P(c)$	.98 (.76)	.90 (.74)	1.04 (.77)
2	d' $P(c)$	.92 (.74)	1.07 (.77)	1.11 (.78)
3	d' $P(c)$	.83 (.72)	.79 (.71)	.85 (.73)
Average	d' $P(c)$	.91 (.74)	.92 (.74)	1.00 (.76)

The results obtained with the 2AFC procedure in Cond. 2, 3, and 4 are shown in Table 3. (Small discrepancies in the correspondence between values of $P(c)$ and d' are attributable to the fact that d' was calculated on the basis of values of $P(c)$ read to three places, although the table shows $P(c)$ rounded to two places.)

The 2AFC procedure clearly shows a smaller effect of the motivational manipulations. The difference in $P(c)$ between ML I and ML II is 0.0; between ML II and ML III, .02. The latter difference is the equivalent of about .4 db. The standard error of $P(c)$ with 800 observations, assuming binomial variance (an underestimation of the actual variance), is about .015; thus the difference does not approach statistical significance.

As an aside, a comparison of Tables 2 and 3 indicates the comparability of estimates of d' obtained under the rating and 2AFC procedures. Under ML II the difference is less than .5 db.; the difference under ML III is obviously of no significance. As a further aside, the three *O*s earned bonuses of $11.40, $16.80, and $9.00 under the rating procedure, and $1.20, $.75, and $.75 under the 2AFC procedure.

DISCUSSION

Four studies reported previously (Blackwell, 1953; Gundy, 1961; Lukaszewski & Elliott, 1962; Zwislocki, Maire, Feldman, & Rubin, 1958) bear directly on the subject at hand.

Gundy determined the index *d′* by the rating method. He examined the effect of practice in a single 1-hr. session which consisted of eight blocks of 50 trials. Three independent variables were employed: feedback vs. no feedback of results, low vs. high signal intensity, and signal specification vs. no signal specification. The last variable mentioned refers to the practice of clearly identifying the signal for *S*s before the test trials begin; in this experiment the signal was presented on three trials before each block of 100 trials at an amplitude 10 db. greater than that used in the experiment.

Gundy found a practice effect of 4–5 db. over the first six blocks of 50 trials when the signal was not specified. When the signal was specified, the practice effect was considerably smaller: approximately 1.5 db. when feedback was given, and less than .5 db. when no feedback was given. Trial-by-trial feedback helped little—by less than 1.0 db.—whether or not the signal was specified. The effect of signal specification in the first block of 50 trials was significant; it aided detection by 7 db. at the low signal intensity, and by 3 db. at the high signal intensity. On the last two blocks of 50 trials, the signal was specified for all *O*s, and any differences evaporated.

Zwislocki et al. (1958) conducted five experiments. For a signal of 100 cps (a frequency known to be more susceptible to a practice effect than higher frequencies), for a combined forced-choice and tracking procedure, and with feedback supplied, a practice effect of 10 db. was observed. Of the 10 db., 6 db. were accomplished in the first 50 trials; an additional 2 db. were accomplished in the next 100 trials; and the remaining 2 db. were accomplished before 450 trials had taken place. For a 100-cps signal, under the 2AFC procedure, with feedback and a bonus, a practice effect of 5 db. was observed over eight blocks of 100 trials. For a 1,000-cps signal, under the Békésy tracking procedure, with no feedback or bonus, a practice effect of 2.5 db. was accomplished in four ½-hr. sessions. The same *O*s then showed a 4-db. effect at 100 cps, and an additional 4-db. effect when feedback and a bonus were introduced.

With a new set of *O*s, under the tracking procedure, without feedback or bonus—and a signal of 100 cps—a 5-db. effect was observed over three ½-hr. sessions. At the seventh session the introduction of a bonus had no effect; a 2-db. effect of the bonus was observed in the eighth session. In another experiment like the preceding except that feedback and a bonus were supplied from the beginning, an effect of 9 db. was observed between Session 1 and Sessions 4–7. The threshold in the first session was 5.5 db. lower than that in the first session of the preceding experiment, a difference which did not reach statistical significance.

Lukaszewski and Elliott (1962) studied the four combinations of two two-state variables: forced-choice vs. non-forced-choice and feedback vs. no feedback. With each of the four procedures there was a practice effect of approximately 2 db. over 7 sessions. In every case subsequent sessions showed some deterioration of performance; for example, for the group working under the forced-choice procedure with feedback, performance in the eighth session was just .75 db. better than in the first session. Trial-by-trial feedback led to thresholds lower by about 3.5 db., from the first session on, than the procedures without feedback. After 12 sessions, the introduction of a bonus produced a 1-db. effect in 2 additional sessions.

Blackwell (1953) conducted a program of experimentation, with visual signals, too extensive for the particulars to be recounted here. In general, with both the forced-choice and yes-no procedures, he found a practice effect over five or six 2-hr. sessions. The *S*s who were given a bonus throughout yielded thresholds which reached a lower asymptote than did the thresholds of *S*s who were not given a bonus. This same asymptote was reached, however, without a bonus, if *S*s received trial-by-trial feedback, or if they simply received their scores after each block of 50 trials. A bonus first given after an asymptote was reached lowered the threshold further if the asymptote were reached without feedback. If the asymptote were reached with feedback, immediate or delayed, the introduction of a bonus did not bring about further improvement.

It can be seen that the present study is in substantial agreement with the studies of Gundy and of Zwislocki et al., with respect to the effect of practice. With *S*s working under a standard level of motivation, no effect was observed over five sessions in the present study. Gundy obtained an effect of less than 1.5 db., which was accomplished in one session. With the forced-choice procedure, Zwislocki obtained a large (5 db.) effect, at a low frequency, but its duration was less than a single session. Lukaszewski and Elliott, and Blackwell, on the other hand, obtained practice effects of longer duration, on the order of six sessions.

Concerning the manipulation of motivation after initial training, three of the four relevant studies agree in showing a small effect, 1 db. or less. In the present study, added motivation had an effect of less than 1 db. with the rating procedure, and an effect of less than .5 db. with the forced-choice procedure. In Blackwell's study, added motivation had no noticeable effect if feedback had been given during training. The effect in the study of Lukaszewski and Elliott was about 1 db. Zwislocki et al., using a tracking procedure, found a 2-db. effect. Thus the data do not support the assumption (Atkinson, 1963) that detection performance will vary from near chance to near perfect as a result of experimental manipulation of motivation.

CONCLUSION

A general conclusion is possible. For audition the effect of practice is limited to the first session (and is no more than 0–2 db. for other than low frequencies), and for audition and vision the effect of added monetary motivation, if there is one, is of little consequence (1 db. or less)—under the forced-choice or rating procedures, with signal specification and feedback.

REFERENCES

Atkinson, R. C. A variable sensitivity theory of signal detection. *Psychol. Rev.*, 1963, **70**, 91–106.

Blackwell, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Bull. Engng. Res. Inst. U. Mich.*, 1953, No. 36.

Egan, J. P., Schulman, A. I., & Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. *J. Acoust. Soc. Amer.*, 1959, **31**, 768–773.

Green, D. M., Birdsall, T. G., & Tanner, W. P., Jr. Signal detection as a function of signal intensity and duration. *J. Acoust. Soc. Amer.*, 1957, **29**, 523–531.

Gundy, R. F. Auditory detection of an unspecified signal. *J. Acoust. Soc. Amer.*, 1961, **33**, 1008–1012.

Lukaszewski, J. S., & Elliott, D. N. Auditory threshold as a function of forced-choice techniques, feedback, and motivation. *J. Acoust. Soc. Amer.*, 1962, **34**, 223–228.

Speeth, S. D., & Mathews, M. V. Sequential effects in the signal-detection situation. *J. Acoust. Soc. Amer.*, 1961, **33**, 1046–1054.

Swets, J. A. Detection theory and psychophysics: A review. *Psychometrika*, 1961, **26**, 49–63.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

Tanner, W. P., Jr., & Swets, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

Zwislocki, J., Maire, F., Feldman, A. S., & Rubin, H. On the effect of practice and motivation on the threshold of audibility. *J. Acoust. Soc. Amer.*, 1958, **30**, 254–262.

(Early publication received April 18, 1963)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 127–132

INFLUENCE OF INSTRUCTIONAL SET AND RESPONSE FREQUENCY ON RETROACTIVE INTERFERENCE[1]

HENRY A. SCHWARTZ[2]

Johns Hopkins University

In an A-B, A-C, A-B retroaction situation, 60 Ss learned an original list of paired associates in which the response members were low frequency associates to the stimulus members. In IL (interpolated learning), the response members were either high or other low frequency associates. In IL Ss were told that learning IL would make OL (original learning) easier (Easy set) or harder (Hard set) to remember. A control group (Neutral set) was told nothing of the possible influence of IL. A Set × IL Response Frequency analysis of variance showed superior recall for the High Frequency IL response groups ($p < .001$) and for the Hard set groups ($p < .001$). No significant interaction of Set with IL Response Frequency was found.

Interference theories of retroaction have generally assumed that the inhibitory effects observable in recall are the result of competition between learned habits of different strengths. Relatively little consideration, however, has been given to the role of more general factors, such as instructional set, in retroaction.

In the standard retroaction experiment, *S*, while learning the original task (OL), is unaware that he will later be required to remember it. Moreover, when the interpolated task (IL) is introduced, *S* is not told of its possible influence on the retention of OL. Lester (1932), however, has presented data which suggest that the amount of interference varies as a result of the set *S* is given at the time of OL. Lester's procedure, however, did not rule out the possibility that the obtained differences in interference were merely due to differing amounts of rehearsal during the 24-hr. OL-RL interval employed.

The present experiment is designed to study the influence of instructional set upon retroactive interference and the possible interaction of set with response frequency, a variable important to response competition theory.

According to Osgood (1953), one of the most obvious principles underlying retroactive interference is response competition, which operates in a manner such that, "Whenever the same stimulus tends to evoke two or more incompatible reactions, that one having the strongest habit strength in association with the stimulus will occur [p. 551]."

Since associative response frequency is considered to be a measure of the strength of an association, the principle of response competition would predict in a retroaction situation where OL and IL stimuli are identical, that greater interference should result from interpolating high frequency, rather than low frequency, associates as IL responses.

---

[1] Based upon a dissertation submitted to the Faculty of Philosophy, the Johns Hopkins University, in partial fullfillment of the requirements for the PhD degree. This investigation was carried out during the tenure of a Predoctoral Fellowship (12, 164) from the National Institutes of Mental Health, United States Public Health Service. The author is deeply indebted to James E. Deese for his guidance throughout the course of this investigation.

[2] Now at International Business Machines Corporation, White Plains, New York.

The present study employs paired-associate tasks in which the OL response members are all low frequency associates to the stimulus members, as determined from normative data. The IL responses are either high frequency, or other low frequency, associates to the same stimulus members used in OL.

METHOD

Material.—The learning materials employed were stimulus-response pairs of the type used by Deese and Hardman (1954) and consisted of short sentences composed of a definite article (in some cases), subject, predicate, and predicate adjective (for example, "The valley is deep."). The first few words constituted the stimulus member and the final word the response member.

To obtain this material, 100 nouns of AA frequency value were selected from the Thorndike-Lorge (1944) word counts. These nouns were then put in incomplete sentence form, such as "The valley is ———," and were presented on mimeographed sheets to 66 Johns Hopkins University undergraduate students enrolled in introductory psychology. These students were instructed to write in the blank space the first adjective they thought of after reading each sentence. From these data, 12 incomplete sentences were selected as stimulus members. For each of these, two low frequency and one high frequency associates were selected as response members. Low frequency responses were defined as those which had occurred in the norms as a response to a particular stimulus only once out of the possible 66 times. High frequency responses were those which had occurred in the norms 20 or more times as responses to the particular stimulus item. The mean frequency of occurrence of the high frequency responses in the norms was 35.50.

The same 12 incomplete sentences were employed as stimulus members in OL and IL. The response member in OL, to each of the stimuli, was one of the low frequency associates to that stimulus. The response member in IL, to each of the stimuli, was either the high frequency or the other low frequency associate to that stimulus.

Subjects.—The *S*s were 60 male Johns Hopkins University undergraduate students from an introductory course in psychology. None of these *S*s had participated in the gathering of the normative data from which the S-R material was selected.

Experimental design.—The experimental design consisted of two levels of IL response frequency combined factorially with three levels of instructional set during IL, yielding a six-cell matrix with 10 *S*s per cell.

Procedure.—The S-R material was presented on a modified memory drum. The stimulus member appeared for 1 sec. and was followed by the stimulus and response members together for 1 sec. The pairs appeared in a random order on each presentation of the list. The interitem interval was 1 sec. and the intertrial interval 2 sec.

All *S*s were given instructions for OL and were then given two trials on an unrelated list to allow them to familiarize themselves with the operations required. This list consisted of names and occupations to be associated, e.g., "Tom is a *Printer.*" The *S*s then learned OL to a criterion of one errorless trial. IL took place immediately after OL. For IL, *S*s were randomly divided into six groups according to the particular combination of the independent variables they were to receive. These *S*s were then given the appropriate instructions, as follows:

Neutral groups:

> Now, before I test you to see how well you can remember what you have just learned, I have another task for you to learn. Again, try to do your best. The procedure and the phrases in this task are the same as in the previous one, however, the completing words have been changed. You are to learn the new words for the phrases just as you learned the previous ones. As before, you should say each phrase and the completing word as they appear on the first presentation of the list. After that, you should try to say the correct completing word when only the phrase appears.

Easy groups—Same as for Neutral groups, plus:

> By now, in your psychology course, you may have studied something about memory. As you probably know, after a list of words has been learned, learning a second list may make the first list either harder or easier to remember, depending upon the words used. There appears to be some kind of unconscious connections between certain words so that learning one can make it easier to remember the other. I am trying to determine just how this occurs.

> I have chosen the completing words in this task so that learning them should make it much easier for you to remember the first list of words. I want to see if you can take advantage of this condition.

Hard groups—Same as for Neutral groups, plus:

> By now, in your psychology course, you may have studied something about memory. As you probably know, after a list of words has been learned, learning a second list may make the first list either harder or easier to remember, depending upon the the words used. I am trying to determine just how this occurs.
>
> I have chosen the completing words in this task so that learning them should make it much harder for you to remember the first list of words. I want to see if you can avoid confusing the words on the two tasks.

The *S*s then learned IL to a criterion of one errorless trial. Ten minutes after OL, *S*s relearned OL to a criterion of one errorless trial. Both the OL-IL and the OL-RL intervals were controlled. Thus, the IL-RL interval inevitably varied; however, studies (Briggs, 1954; Underwood, 1948) have shown interference effects to be relatively independent of variations in the length of this interval.

To control for rehearsal and warm-up decrement the IL-RL interval was filled by having *S*s practice the familiarization task.

RESULTS

Table 1 shows the mean performance of each group for original learning, interpolated learning, relearning, and recall.

A between-within analysis of variance was performed on the trials to criterion data for OL. As would be expected, the obtained *F* was nonsignificant ($p > .05$). Therefore, the results subsequently obtained in IL, RL, and recall are not attributable to differences in original learning.

Having found no significant difference between the groups on OL, the data for IL, RL, and recall were each subjected to a two-way analysis of variance (Set × IL Response Frequency).

In IL, the high frequency responses were learned in fewer trials than the low frequency responses, $F(1, 54) = 21.07$, $p < .001$. The greater ease of learning of the high frequency pairs appears to justify the assumption that the normative frequencies reflect strength of association. Neither the effects of set nor those of the Set × Response Frequency interaction were significant ($p > .05$). Thus, it appears that *S*s' performances on IL were not influenced by the instructions regarding the effect of such learning upon the retention of the original task.

TABLE 1

MEANS AND *SD*s OF NUMBER OF TRIALS TO CRITERION FOR ORIGINAL LEARNING, INTERPOLATED LEARNING, AND RELEARNING, AND MEANS AND *SD*s OF NUMBER OF WORDS CORRECTLY ANTICIPATED IN RECALL

Group		Task							
		OL		IL		RL		Recall	
IL Response Frequency	IL Set	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
High	Hard	8.10	3.88	3.80	2.14	5.60	3.75	8.10	1.51
High	Easy	7.30	3.49	4.50	2.25	5.10	1.92	5.40	1.50
High	Neutral	7.30	2.33	3.10	1.58	5.80	2.99	5.20	1.08
Low	Hard	7.40	2.87	6.80	2.82	5.80	3.12	5.70	1.42
Low	Easy	6.10	3.70	7.00	2.19	5.30	1.90	4.60	1.28
Low	Neutral	7.20	3.54	6.90	3.48	6.80	3.49	4.20	1.60

TABLE 2

Analysis of Variance of Recall Scores

Source	*df*	*MS*
Set (C)	2	28.47***
Response frequency (R)	1	13.37***
C × R	2	3.80
Error (*S*s within cells)	54	(2.20)

*** $p < .001$.

In RL, the number of trials required to reach criterion was not significantly affected by either set or IL response frequency, nor was there any significant interaction ($p > .05$). This is not surprising, since relearning is known to be a rather insensitive measure of interference effects.

When the data for recall (number of words correctly anticipated on the first RL trial) were analyzed, the values shown in Table 2 were obtained.

As shown in Table 2, both main effects were highly significant in recall. Interpolating another low frequency response produced more interference than that produced by interpolating a high frequency response. Also, instructing *S*s that IL would interfere with retention resulted in greater recall then either the "Easy" or the "Neutral" instructions. Tukey's (in Ryan, 1959) method for multiple comparisons, performed for the three levels of set, showed that the "Hard" set differed significantly from either of the other two levels of set ($p < .001$). No significant difference was found between the "Easy" and the "Neutral" sets ($p > .05$).

The interaction in recall, as in the IL and RL analyses, was nonsignificant ($p > .05$), indicating that the effects of response frequency were approximately the same for each of the three levels of set.

Tables 3 and 4 present summaries of the types of errors made, respectively, in IL and recall. Since the number of IL trials varied for the groups, the values shown in Table 4 represent errors per trial excluding the IL criterion trial. Due to their theoretical importance in interference studies, the data for interlist errors in Tables 3 and 4 were subjected to the Set × IL Response Frequency analysis of variance. No significant differences in the number of interlist errors per trial were found in either the IL or the recall analyses ($p > .05$).

All *S*s were interrogated immediately following RL to try to get some indication of the techniques they

TABLE 3

Means and *SD*s of Errors per *S* per Trial by Category for IL

Group		Type of Error							
		Interlist		Intralist		Extralist		Response Omission	
IL Response Frequency	IL Set	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
High	Hard	.32	.35	.24	.40	.10	.30	1.81	1.54
High	Easy	.29	.39	.16	.30	.13	.22	2.05	1.25
High	Neutral	.26	.45	.14	.30	.00	.00	1.50	1.25
Low	Hard	.05	.14	.37	.30	.05	.10	2.79	.82
Low	Easy	.31	.26	.38	.35	.06	.14	2.79	1.08
Low	Neutral	.47	.71	.58	.66	.22	.24	2.70	1.38

Note.—Criterion trial omitted.

TABLE 4

MEANS AND *SD*s OF ERRORS PER *S* BY CATEGORY FOR RECALL

Group		Type of Error							
		Interlist		Intralist		Extralist		Response Omission	
IL Response Frequency	IL Set	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
High	Hard	1.10	1.14	1.00	1.10	0.00	0.00	1.80	1.17
High	Easy	1.80	1.25	1.00	1.26	0.10	0.30	3.70	1.68
High	Neutral	1.00	0.78	2.00	1.18	0.10	0.30	3.70	1.68
Low	Hard	0.70	1.27	0.90	1.04	0.30	0.64	4.40	1.43
Low	Easy	0.60	0.66	1.00	0.63	0.30	0.64	5.50	1.57
Low	Neutral	1.40	1.36	1.80	1.78	0.40	0.49	4.20	1.99

employed in learning and remembering the tasks. To conserve space and because *S*s' answers generally failed to reveal any patterns which might differentiate the groups, these data are not presented here. The interrogation, however, did reveal that in the case of the High Frequency groups, the difference between the OL and IL responses was clearly noted. Each of these *S*s reported that the IL responses seemed more appropriate than the OL responses. Oddly, of the Low Frequency groups, 12 *S*s reported the responses in either OL or IL as being more appropriate.

DISCUSSION

The effects of IL response frequency do not appear to lend themselves to interpretation within the framework of the response competition principle stated by Osgood (1953, p. 551). There appear to be at least three likely possibilities. First of all, the greater number of IL trials taken by the Low Frequency groups allowed increased opportunity for the unlearning of the OL responses. Since the frequency of overt interlist intrusions during IL is considered to be a direct index of the unlearning factor (Osgood, 1948), an analysis was made of the total number of interlist errors occurring in IL, disregarding the differing numbers of trials taken. No significant differences were found. Thus, by this index, unlearning cannot account for the results of the experiment. Secondly, it is possible that the high frequency IL responses were more readily rejected during recall because of higher accuracy in identification of list membership. This viewpoint, however, would have predicted that the High Frequency groups would make fewer overt interlist errors during recall, but no such difference was found. Finally, the rate of dissipation of interference effects might have been greater for the High Frequency groups. Interference effects have, in fact, been found to dissipate more rapidly with higher degrees of IL (Melton & Irwin, 1940).

The effects of instructional set make it evident that the amount of interference which may occur in a situation is not fixed primarily by the classical variables, such as type of material, time intervals, etc., but is also influenced by the set *S* is given. The effects of set upon recall are interesting when coupled with the fact that no significant difference due to this variable was found in IL. It appears that *S*s, when warned of the interfering effects of IL upon retention, can learn IL just as well as *S*s not given such information, but can retain more of OL while doing so. The determination of the mechanisms through which instructional set operates, however, awaits future investigation.

REFERENCES

BRIGGS, G. E. Acquisition, extinction, and spontaneous recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

DEESE, J., & HARDMAN, G. W., JR. An analysis of errors in retroactive inhibition of rote verbal learning. *Amer. J. Psychol.*, 1954, **67**, 299–307.

LESTER, O. P. Mental set in relation to retroactive inhibition. *J. exp. Psychol.*, 1932, **15**, 681–699.

MELTON, A. W., & IRWIN, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

OSGOOD, C. E. An investigation into the causes of retroactive interference. *J. exp. Psychol.*, 1948, **38**, 132–154.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

RYAN, R. A. Multiple comparisons in psychological research. *Psychol. Bull.*, 1959, **56**, 26–47.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

UNDERWOOD, B. J. Retroactive and proactive inhibition after 5 and 48 hours. *J. exp. Psychol.*, 1948, **38**, 29–38.

(Received July 5, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 133–137

DEVELOPMENTAL CHANGE IN THE RELATIVE VALUES OF SOCIAL AND NONSOCIAL REINFORCEMENT

MICHAEL LEWIS,[1] A. MARTIN WALL,[2] AND JUSTIN ARONFREED

University of Pennsylvania

Asymptotic levels of performance in a binary choice paradigm were used to assess developmental change in the relative reinforcement values of social approval and impersonal stimulus events which merely indicated the correctness of prediction of an uncertain outcome. Ss were 30 1st graders and 30 6th graders, randomly assigned in equal numbers to each of the reinforcement conditions. Only the 1st graders under social reinforcement tended toward maximizing the more frequently reinforced response, while the remaining 3 groups showed asymptotic proportions of response below the actual probability of being correct. It was suggested that the value of repeated external social approval decreases with age and that the intrinsic reinforcement of correct prediction is greater in older children. The interpretation of the findings also emphasized the minimal role of mediating concepts of probability and the possible effects of having used contingent outcomes.

Various theoretical treatments of internalized self-reinforcing patterns of behavior (Aronfreed, 1964; Hill, 1960; Mowrer, 1960, Ch. 10; Sears, 1957) have noted the initial dependence of young children upon direct social reinforcement and their subsequent tendency to rely increasingly, as their social experience accumulates, on stimulus events more intrinsically related to their own responses. Very little experimental work has been done, however, to establish the relative effectiveness of qualitatively different kinds of reinforcing events at varying points in the course of development. The study reported in this paper uses the binary choice learning paradigm to assess the values of two types of reinforcement for children of two distinct age groups. The value of social reinforcement, delivered entirely in the social behavior of another person, is compared to the value of nonsocial, impersonal stimulus events which simply indicate correctness of the child's prediction of an uncertain outcome. The purpose of the study is to illustrate how probability learning may be employed as a least common denominator for evaluating the reinforcing properties of qualitatively different events and for detecting changes in the relative value of such events.

Statistical learning models (Bush & Mosteller, 1955; Estes, 1957) originally tended to emphasize the conditions under which *S*s might match the frequencies of their responses to the frequencies of two events predicted by the responses, when these events occurred with fixed probabilities in a random sequence. While such matching behavior has often been observed (Estes & Straughan, 1954; Grant, Hake, & Hornseth, 1951; Humphreys, 1939), *S*s may also maximize the correctness of their responses by consistently predicting the more frequently occurring event. Recent learning and decision-making models have suggested that asymptotic levels of response are a function of the magnitude of explicit "payoffs" beyond mere correctness of prediction (Bush & Estes, 1959; Edwards, 1954; Siegel, 1959), and various experiments have demonstrated maximum gain responding when outcomes incorporate concrete external reinforcements (Bush & Mosteller, 1955, Ch. 13; Edwards, 1956; Galanter & Smith, 1958; Meyer & Seidman, 1960; Siegel & Goldstein, 1959). Pubols (1960) has recently summarized evidence pointing to the effect of incentive value upon asymptotic level of performance.

---

[1] Now at the Fels Research Institute for the Study of Human Development, Yellow Springs, Ohio.

[2] Now at the Department of Psychology, University of Toronto.

In the present experiment, asymptotic frequency of the response that had the greater probability of being correct was used as an index of reinforcement value. Verbal approval, a social reinforcement, was expected to have greater value for first-grade children than the nonsocial reinforcement of a light that signified a correct response. However, no difference in favor of social reinforcement was expected among sixth-grade children. These expectations were based on the view that repeated social approval would show a decrement in value with increasing age (because of changes in the strength of relevant motivational dispositions) and that the intrinsic reinforcement attached to being correct would show a corresponding increment. Presumably, first-grade children would have internalized values directed toward being consistently correct. But the use of social reinforcement was expected to result in their maximizing the most frequently correct response because the value of the approval would override the value of predicting an uncertain event. In contrast, sixth graders were expected not to drift toward maximizing approval, since the value of exact prediction would be greater than the value of social reinforcement for them.

METHOD

Subjects.—The *S*s were 60 boys drawn in equal numbers from the first and sixth grades of a public school. In each grade, 15 *S*s were randomly assigned to each of the two reinforcement conditions.

Apparatus.—The apparatus consisted primarily of two metal boxes, approximately 5 × 8 × 12 in. in size. The *S*'s box contained a large green light on the front panel. Below and to either side of the light were two lever switches. Within the box was a buzzer used to signal the onset of each trial. A long cable, not visible to *S*, connected this first box to a second *E*'s box in an adjacent room. The second box had lights which indicated *S*'s choices and switches which operated the green light and the buzzer in *S*'s box.

Procedure.—The basic procedure of the experiment was the same for conditions of both social and nonsocial reinforcement. The *S* was aware of the presence of only one *E* (E_1), who was the same individual under all conditions. The E_1 told *S* that he would be playing a game in which he was to push down only one of the two levers each time that he heard the buzzer from within the machine. The instructions conveyed that the object of the game was to try to produce the reinforcing event. It was intended that *S* not be able to infer, in the absence of reinforcement, that the alternative response was correct, so that *S*'s knowledge of the correct alternative would be contingent on his choice.

The *S* was instructed to press the left lever on the first two of four sample trials and the right lever on the last two trials. Reinforcement was given on the first and fourth trials. Then *S* began the series of 150 training trials. Events were randomized within each block of 25 trials, with the overall probabilities of the left and right levers being correct in the ratio of .70:.30. The intertrial interval was roughly 3 sec.

The second *E* (E_2) used the control box in the adjacent room and recorded *S*'s responses. In the nonsocial reinforcement condition, E_2 also controlled presentation of the green light on *S*'s box for the duration of about 1 sec., when *S*'s response was correct.

Under nonsocial reinforcement, E_1 pointed out, after the sample trials, that the machine would itself be able to indicate whether *S*'s responses were correct. Then E_1 left the room after explaining that he had other things to do. This procedure gave the light an entirely nonsocial significance, since *S* could not even be reinforced by social observation of his correct responses.

In the social reinforcement condition, E_1 remained seated just behind *S* throughout the entire series of trials and said either "Good!" or "Fine!" whenever *S*'s response was correct. The green light was not used.

RESULTS AND DISCUSSION

Figure 1 shows, for both age groups, mean proportions of responses to the more frequently reinforced side ($S_{.70}$), in successive blocks of 10 trials, under social (S) and nonsocial (NS) reinforcement. The performance curve for first-grade children who received social reinforcement is clearly distinct from the curves for the other groups. A comparison of the four groups, using the Kruskal-Wallis test (Siegel, 1956) applied to individual *S*s' overall mean proportions of $S_{.70}$ responses, across the entire 150 trials, indicated that these values were not drawn from the same population ($p < .001$). Mann-Whitney *U* tests (Siegel, 1956) were applied to the differences between each two of the groups in their asymptotic proportions of the $S_{.70}$ response over the last two blocks of trials. Asymptotic values for first-grade children who were given social reinforcement were significantly higher than values for any of the other three groups ($p < .001$ in each case). There were no significant differences among these remaining three groups.

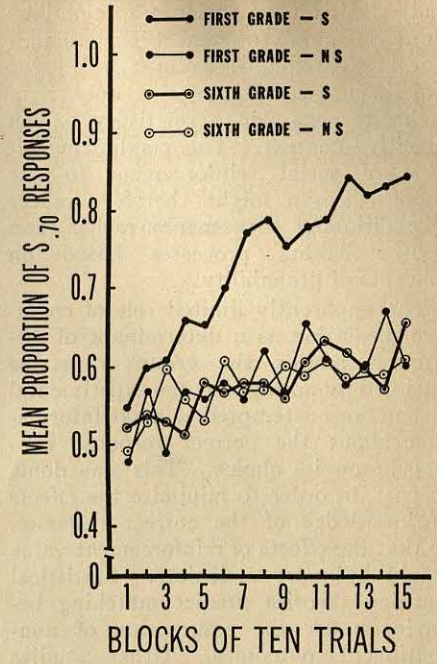


FIG. 1. Mean proportions of $S_{.70}$ responses in successive 10-trial blocks for four experimental groups varying in age and type of reinforcement for correct response (each plotted point is a Group × Block mean based on 15 cases).

These findings confirm the expectation that the greater value of social than of nonsocial reinforcement for first graders would not have a parallel among sixth graders. They are in substantial agreement with the findings of other studies of binary choice behavior. When outcomes are primarily defined by correct prediction, children give the more frequently reinforced response with a probability equal to or less than the probability of its being correct (Kessen & Kessen, 1961; Messick & Solley, 1957). But they show more of a maximizing tendency when additional reinforcement of significant value to them is associated with being correct (Brackbill, Kappy, & Starr, 1962; Meyer & Seidman, 1960; Siegel & Andrews, 1962).

The response patterns of the various groups cannot easily be attributed to cognitive strategies. Piaget (1950) and others (Crandall, Solomon, & Kellaway, 1961; Gratch, 1959) have reported evidence of an obvious increase with age of the comprehension of quantitative properties of uncertain events. But Siegel and Andrews (1962) have asserted that the maximizing behavior of young children, under conditions of high "payoff," implies rudimentary concepts of probability at ages earlier than those which Piaget indicated. However, Piaget's techniques demand a verbal and cognitive evaluation of probability from the child, whereas binary choice learning paradigms, as in the present experiment, usually require only nonverbal responses in a relatively barren cognitive context. The maximizing effects of social reinforcement in the younger group might therefore reflect a conditioning phenomenon rather than decision-making processes based on concepts of probability.

The apparently limited role of cognitive mediation as a determinant of differences among the groups must be qualified by noting that the experimental instructions attempted to make information about the correct response contingent on its choice. This was done, in part, in order to minimize the effects of knowledge of the correct response, so that the effects of reinforcement value would be more articulate. Statistical learning theories predict matching behavior under the assumption of noncontingent outcomes. Other studies (Brackbill et al., 1962; Siegel & Andrews, 1962) which have used noncontingent outcomes with children have also found however, effects of magnitude of reinforcement which are like those reported here. Nevertheless, age differences more sensitive to variations in cognitive approach might have been obtained had the children been able to read the absence of reinforcement with some certainty as an indication that the alternative response was correct.

REFERENCES

ARONFREED, J. The origin of self-criticism. *Psychol. Rev.*, 1964, in press.

BRACKBILL, Y., KAPPY, M. S., & STARR, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **53**, 32–35.

BUSH, R. R., & ESTES, W. K. (Eds.) *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

CRANDALL, V. J., SOLOMON, D., & KELLAWAY, R. A comparison of the patterned and nonpatterned probability learning of adolescent and early grade school-age children. *J. genet. Psychol.*, 1961, **99**, 29–39.

EDWARDS, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380–417.

EDWARDS, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GALANTER, E. H., & SMITH, W. A. S. Some experiments on a simple thought-problem. *Amer. J. Psychol.*, 1958, **71**, 359–366.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

GRATCH, G. The development of the expectation of the non-independence of random events in children. *Child Develpm.*, 1959, **30**, 217–227.

HILL, W. F. Learning theory and the acquisition of values. *Psychol. Rev.*, 1960, **67**, 317–331.

HUMPHREYS, L. G. Acquisition and extinction of verbal expectation in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

KESSEN, W., & KESSEN, M. L. Behavior of young children in a two-choice guessing problem. *Child Develpm.*, 1961, **32**, 779–788.

MESSICK, S. J., & SOLLEY, C. M. Probability learning in children: Some exploratory studies. *J. genet. Psychol.*, 1957, **90**, 23–32.

MEYER, W., & SEIDMAN, S. B. Age differences in the effectiveness of different reinforcement combinations on the acquisition and extinction of a single concept learning problem. *Child Develpm.*, 1960, **31**, 419–429.

MOWRER, O. H. *Learning theory and the symbolic processes.* New York: Wiley, 1960.

PIAGET, J. Une expérience sur la psychologie du hasard chez l'enfant: Le tirage au sort des couples. [An experiment in the psychology of chance in the child: Drawing pairs.] *Acta psychol.*, 1950, **7**, 323–336.

PUBOLS, B. H., JR. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

SEARS, R. R. Identification as a form of behavioral development. In D. B. Harris

(Ed.), *The concept of development.* Minneapolis: Univer. Minnesota Press, 1957. Pp. 149–161.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SIEGEL, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

SIEGEL, S., & ANDREWS, J. M. Magnitude of reinforcement and choice behavior in children. *J. exp. Psychol.*, 1962, **63**, 337–341.

SIEGEL, S., & GOLDSTEIN, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

(Received August 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 138–141

RESPONSE SHAPING AT LONG INTERSTIMULUS INTERVALS IN CLASSICAL EYELID CONDITIONING[1]

WILLIAM F. PROKASY, HARVEY C. EBEL, AND DONALD D. THOMPSON

Pennsylvania State University

The performances of 3 groups of human Ss in a classical eyelid conditioning situation were compared at an interstimulus interval (ISI) of 2.497 sec. 1 group received all training at an ISI of 2.497, a 2nd group received prior training at an ISI of .63 and was switched to one of 2.497, while the 3rd group began training at .63 and was gradually changed to one of 2.497 in 6 steps. The frequency of response was greater in the gradually shifted group than in the other 2 groups. It was concluded (a) that whether or not high response frequencies were obtained at long ISIs depends upon the specific training regimen; (b) that a time discrimination, in the form of response latency, played an important role in the results; and (c) that asymptotic performance at one ISI is not independent of experience with other ISIs.

The degree to which latency of the classically conditioned eyelid response (CR) can be modified by manipulations of the interstimulus interval (ISI) (Boneau, 1958; Ebel & Prokasy, 1963) indicates, as Spence (1956) has suggested, that response latency is important in the acquisition of the eyelid CR.

The purpose of this study is to examine further the role of latency in the acquisition of conditioned eyelid responses. It is predicated on two general beliefs: (*a*) that reinforcement is necessary for CR performance and (*b*) that the efficiency of CR latency with respect to the UCS determines, in part, the degree of reinforcement. According to this view, if *S*s can learn to make an appropriate time discrimination, then the low performance usually obtained at long ISIs (e.g., Kimble, 1961, pp. 156, 157) will be improved. This study constitutes an attempt to "shape" a more appropriate time discrimination by initiating training at an optimum ISI and gradually lengthening it to one at which performance is normally quite low.

METHOD

Apparatus.—The apparatus for picking up, amplifying, and recording eyelid responses is described elsewhere (Prokasy & Whaley, 1961).

The CS was an increase in brightness from 5.2 to 295.6 mL. of a 3-in. milk glass disk located approximately 32 in. in front of *S*, about 11 in. below eye level. The UCS was a 50-msec. puff of nitrogen delivered to *S*'s right cornea through a 4.5-mm. diameter jet adjusted to between .25 and .75 in. from *S*'s eyelash. The puff intensity was sufficient to support a 320-mm. column of mercury, measured at point of delivery to the eye.

Trials were programed at intertrial intervals of 8, 12, and 16 sec., with an average of 12, through a Western Union tape transmitter. The CS, UCS, and ISI durations were controlled by Tektronix timers. Equipment cues were masked by a 70-db. (re 0002 dynes/cm^2) white noise sound screen presented through a speaker mounted in *S*'s cubicle.

Subjects.—The *S*s were 107 volunteers from

---

[1] This study was supported by funds to the senior author from National Science Foundation Grant G-18119. It was conducted while H. C. Ebel was a Predoctoral National Institute of Mental Health Research Fellow and D. D. Thompson was a participant in a program of Research Participation for College Teachers, supported by the National Science Foundation.

The authors would like to acknowledge the aid of Stanley Lipkin, who participated in the National Science Foundation's Undergraduate Research Participation program.

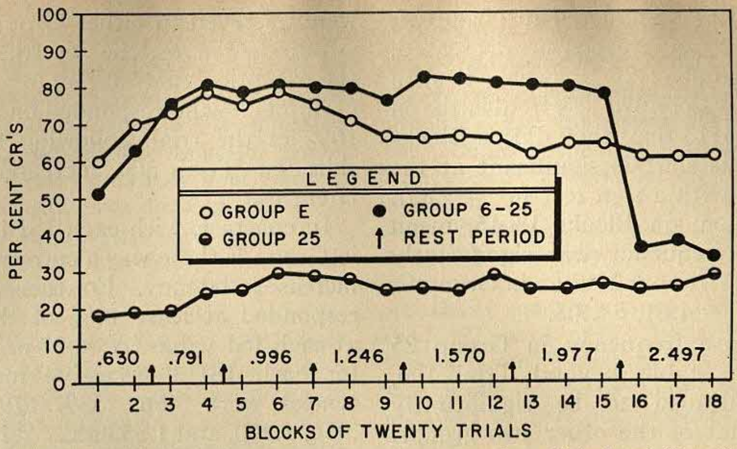


FIG. 1. Percentage of conditioned responses as a function of blocks of 20 training trials. (The arrows indicate the points at which brief rest periods of 30–50 sec. were introduced, while the numerical values between the arrows indicate ISI, in seconds, for Group E.)

introductory psychology courses at the Pennsylvania State University who participated in the study to fulfill a course requirement. Of the 107 *S*s 8 were lost. Two were lost because they adapted to the UCS, 1 because his spontaneous blink rate reached levels as high as six per second over spans of several seconds, 1 because she fell asleep during the conditioning procedure, and the rest for *E* error or equipment failure.

Experimental design.—The *S*s were assigned to one of three treatment conditions. The experimental group (Group E) was given a total of 360 conditioning trials, with 100% reinforcement, in the following order: 40 with ISI = .630 sec.; 40 with ISI = .791 sec.; 50 with ISI = .996 sec.; 50 with ISI = 1.246 sec.; 60 with ISI = 1.570 sec.; 60 with ISI = 1.977 sec.; and 60 with ISI = 2.497 sec. Control Group 25 received a total of 360 trials with ISI = 2.497 sec., while Control Group 6-25 received 300 trials with ISI = .630 sec. followed by 60 with ISI = 2.497 sec. A time break of from 30 to 50 sec. was taken during each of the six ISI shifts in Group E, permitting sufficient time for *E* to change timer settings and for *S* to "stretch and relax." Similar time breaks were introduced at corresponding stages of training in Groups 6-25 and 25. All timer settings were adjusted to allow for a .03-sec. UCS delivery delay, and, for all ISI values, CS duration equaled the ISI.

The *S*s were instructed to sit quietly, not to try to figure out the experiment, and to let their reactions take care of themselves.

CR definition.—The CRs were defined as the occurrence of a 1-mm. or greater pen deflection at least .17 sec. after CS onset, but prior to and overlapping the UCS. There were three reasons for selecting this measure. First, and consistent with our "shaping" assumption, it gave the CR the "functional" property of producing partial UCS avoidance. Second, the opportunity for spontaneous blinks to be recorded as CRs increases with the ISI and can, therefore, produce a bias in the recording. By restricting the measure only to those responses which overlapped the UCS, the likelihood of recording a spontaneous blink as a CR was more nearly equal for the different ISI conditions. Third, the data of Boneau (1958) indicate that CRs tend, as training progresses, more and more to overlap the UCS, thus giving empirical rationale in terms of what *S*s do in the eyelid conditioning situation.

RESULTS

Frequency.—Percent CR as a function of blocks of 20 training trials is plotted in Fig. 1. As anticipated, Groups E and 6-25 (having been exposed to identical ISIs) did not differ significantly, *F* (1, 64) = .91, on the first two training blocks. The effects of the ISI variable are clearly seen on the first two training blocks, as Group 25 responded significantly less than did Group 6-25, *F* (1, 64) = 85.64, and Group E, *F* (1 ,64) = 45.07.

As ISI was increased in Group E, frequency of response decreased. Eighteen out of 26 *S*s (seven ties) in Group E responded less frequently on Blocks 13–15 than they did on Blocks 3–5. This shift is significant at the .05 level with a sign test ($\chi^2 = 3.85$). In addition, on Blocks 13–15, mean response frequency was significantly higher in Group 6-25 than in Group E, $F\,(1, 64) = 4.61, p<.05$.

Response frequency in Group 25 remained stable beyond Trial 100, and continued to be significantly below that of the other two groups. Of specific concern are the group comparisons on Blocks 13–15, prior to shifting Groups E and 6-25 to an ISI of 2.497 sec. Mean frequency was significantly lower in Group 25 than in either Group 6-25, $F\,(1, 64) = 74.89$, or Group E, $F\,(1,64) = 10.42$.

Between Blocks 15 and 16, Group 6-25 was shifted from an ISI of .630 to one of 2.497 sec. This resulted in an immediate drop in response frequency, following which no later increases were observed. After the shift, the response frequency of Group 6-25 across Blocks 16–18 did not differ significantly from that of Group 25, $F\,(1, 64) = 2.11$, but was significantly lower than that of Group E, $F\,(1, 64) = 7.52$. Group E continued to respond significantly more than did Group 25, $F\,(1, 64) = 20.2$.

Latency.—Mean latency for Groups E and 6-25 combined across Blocks 1 and 2 was .477 sec., while mean latency for Group 25 across the first five blocks was 2.307 sec. These means are based only on those *S*s who gave at least four CRs during the span of those trials, the *N*s for Groups 25, 6-25, and E being, respectively, 25, 27, and 28. The shortest latency in Group 25 was longer than the longest latency in either Group E or Group 6-25. No systematic latency shifts were observed in either Group 25 or Group 6-25 across the 360 trials, with the exception that all *S*s ($N = 22$) in Group 6-25 who responded on at least 10% of the trials following the ISI shift on Trial 300 increased their latencies.

In contrast, with each ISI increase in Group E there was a corresponding increase in latency. For those *S*s who responded at least 10% of the time at each ISI value ($N = 27$ or greater for each ISI), successive mean latencies were .456, .539, 679, .802, .956, 1.200, and 1.553 sec. With each of the six shifts in ISI, the number of *S*s increasing, as opposed to decreasing, response latency was significant, by a sign test, beyond the .01 level. The obtained values of χ^2 associated with successive ISI shifts were, respectively, 7.0, 19.2, 9.3, 19.2, 16.3, and 19.6

Average latency across the final three blocks of trials (when all groups were given a 2.497-sec. ISI) for Groups E, 6-25, and 25 were, respectively, 1.553, 1.855, and 2.212 sec. Each mean is based only on those *S*s who responded on at least 10% of the final 60 trials, the respective *N*s being 31, 22, and 25. Average latency for Group E differed significantly from that of Group 25, $F\,(1, 54) = 18.56$, but not from that of Group 6-25, $F\,(1, 51) = 3.03$. The means of Groups 25 and 6-25 were also significantly different, $F\,(1, 45) = 9.18$.

DISCUSSION

This experiment illustrates that whether or not high response frequencies are obtained with long ISIs depends upon the specific training regimen to which *S* is exposed. It is apparent, too, that latency plays an important role in obtaining high frequencies at long ISIs, since the latency of Group E systematically increased with each increase in ISI but was, at the end of training, still substantially shorter than that of Group 25. To the extent that a learned latency, or time discrimination, is a requisite of CR acquisition, present results parallel a classical observation in discrimination training: a difficult discrimination can be acquired much more easily when training begins with an easy discrimination and gradually becomes more difficult.

The relatively high performance levels of all three groups at an ISI of 2.497 sec. contrasts, in particular, with the results of McAllister (1953). McAllister found (*a*) little or no conditioning at a 2.5-sec. ISI and (*b*) virtual extinction in a group shifted from an optimum ISI to one of 2.5 sec. Most likely, this difference in results is attributable to either or both the higher UCS intensity and longer training session of the present study. In any event, our results establish that eyelid conditioning can occur at an ISI of 2.5 sec. under the usual training procedures, and that an extreme shift from optimum ISI to 2.5 sec. does not necessarily result in extinction.

Present results raise a question about the degree to which performance with one ISI is independent of performance under a previously administered ISI. Boneau (1958), Kimble (1961, p. 159), and Ebel and Prokasy (1963) reached the tentative conclusion that, when the ISI is changed during the course of conditioning, the resulting asymptotic performance at the new ISI is unaffected by experience with prior ISIs. The frequency and latency differences between the three groups over the last 60 training trials indicates that this conclusion has limited generality.

Our results also raise a question concerning the interpretation of the ISI function. Beginning with an "optimum" ISI of .4 or .5 sec., performance typically decreases rapidly as ISI either increases or decreases. On the basis of our data we would suggest that the usual ISI function cannot be interpreted to reflect anything more than the relative difficulty of inducing performance under a somewhat restricted set of empirical operations (i.e., fixed ISI). If, on the other hand, one attends to the problem of the levels of performance that can be produced at different ISIs regardless of specific training procedure, then Group E provides some interesting information. The total performance drop of this group from a peak with a .630-sec. ISI to a minimum with a 2.497-sec. ISI was under 20 percentage points. This indicates that, under optimal training circumstances, the difference in response probability associated with ISIs of .630 and 2.497 sec. is small. In addition, the data of Group E suggest that relatively high levels of performance might be sustained at ISIs substantially beyond 2.5 sec.

A final point deserves comment. It is evident that only a quite crude time discrimination is actually necessary for producing partial UCS avoidance. For example, a response style which would "work" is lid closure at CS onset. Judging from the systematic latency shifts in Group E, however, the actual performance is more complicated than this. Furthermore, the "simpler" response style does not appear either in Group 25 or in Group 6-25. In brief, *S*s do more than simply respond to a CS. Why this is the case would appear to merit further empirical and theoretical exploration.

REFERENCES

Boneau, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *J. exp. Psychol.*, 1958, 56, 464–471.

Ebel, H. C., & Prokasy, W. F. Classical eyelid conditioning as a function of sustained and shifted interstimulus intervals. *J. exp. Psychol.*, 1963, 65, 52–58.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* (2nd ed.) New York: Appleton-Century-Crofts, 1961.

McAllister, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, 45, 417–422.

Prokasy, W. F., & Whaley, F. L. The intertrial interval in classical conditioning. *J. exp. Psychol.*, 1961, 62, 560–564.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received August 27, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 142–148

CONDITIONED FRUSTRATION AS A LEARNED DRIVE[1]

ALLAN R. WAGNER

Yale University

Rats were tested in either (a) a stabilimeter in which a distinctive cue was presented prior to a startle producing loud tone, or (b) a hurdle box in which hurdle crossing terminated the same distinctive cue. Experimental Ss which had previously experienced the cue paired with frustrative-nonreward showed, as compared to control Ss, (a) more vigorous startle responses in the stabilimeter, and (b) increasingly faster hurdle crossing over test trials. The apparent energizing and reinforcing effects associated with the cue for experimental Ss were interpreted as supporting the position that stimuli associated with frustrative-nonreward may come to elicit a learned drive of conditioned frustration.

Recently, a number of investigators (e.g., Goodrich, 1959; Spence, 1960, Ch. 6; Wagner, 1961) have followed Amsel (1958) in positing a learned drive of conditioned frustration. *Frustrative-nonreward*, i.e., the nonreinforcement of a previously rewarded appetitional response, is proposed to elicit a primary motivational response of frustration, which in turn is assumed to be conditionable to formerly neutral cues. Theoretical interpretations employing this assumption have enjoyed considerable success (e.g., Amsel, 1958; Barry, Wagner, & Miller, 1962; Goodrich, 1959; Spence, 1960, Ch. 6; Wagner, 1961, 1963) in deriving findings from simple learning situations involving nonreward experience. To date, however, little attempt has been made to evaluate the assumption itself in a reasonably direct manner.

The demonstration of a particular secondary or learned drive involves the demonstration that formerly neutral cues, under the specified circumstances, can acquire certain of the behavioral effects which distinguish primary drives. Brown (1961, Ch. 5), for example, has pointed out, within the context of Hullian theory, that the presence of such cues should energize existing response tendencies and their removal should serve as a reinforcement.

Thus, *fear* as a learned drive was demonstrated by showing that the presence and the removal, respectively, of cues previously associated with painful stimulation, could produce energizing and reinforcing effects. Brown, Kalish, and Farber (1951) showed that the vigor of the unconditioned startle response to a sudden, loud noise was increased by the presence of cues previously paired with electric shock. In addition, Miller (1948) and Brown and Jacobs (1949) showed that escape from cues previously paired with shock provided an adequate reinforcement for the learning of a new response such as lever pressing or hurdle crossing.

The present investigation was designed to evaluate the proposal that cues associated with frustrative nonreward may also come to elicit a learned drive. Conditioned frustration has been compared to fear in many aspects of detail (Amsel, 1958;

---

[1] The investigation was supported by National Science Foundation Grant G13080, and was reported, in part, at the 1961 meeting of the American Psychological Association, in a Symposium on "The Current Status of the Concept of Inhibition."

Wagner, 1963), both being considered learned, aversive, emotional responses. Consequently, the techniques employed were closely patterned after those which had proved useful in the above mentioned investigations of fear. In separate test situations it was attempted to determine, (*a*) whether the presence of cues previously paired with frustrative-nonreward would potentiate the unconditioned startle response, and (*b*) whether the cessation of such cues would serve as reinforcement for a simple hurdle crossing response.

METHOD

Subjects.—The *S*s were 50, experimentally naive, male albino rats purchased from the Charles River Breeding Laboratory. Their ages ranged from 100 to 110 days at the beginning of the experiment.

Experimental design.—Prior to training *S*s were divided into 25 matched pairs according to weight. One member of each pair was randomly designated as an experimental *S* and given paired presentations of a distinctive cue (CS) and nonreward, interspersed among food rewarded trials, in a simple runway. The other member of each pair was run as a yoked control *S*, receiving the same degree of exposure to the CS, and the same experience in the runway with nonreward as its mate, but under the condition that the two events were never paired.

Fifteen of the yoked pairs were randomly assigned for subsequent testing in a hurdle box in which hurdle crossing terminated the CS. The remaining 10 pairs were assigned for testing in a stabilimeter in which the CS was presented prior to a startle-producing noise. Four pairs were eliminated from testing, due to failure of one member of each to complete the training phase, leaving 9 pairs in the stabilimeter and 12 in the hurdle box from which data were collected.[2]

Apparatus

The several experimental components were enclosed, when in use, in the same 7-ft.^{3} refrigerator shell. This arrangement was designed to reduce extraneous stimulation and to maximize the similarity of the CS when presented in the various situations.

Training.—The training apparatus consisted of an enclosed, U shaped alley and a circular retaining cage. The alley was 3½ in. wide and 4 in. high throughout, and was divided by guillotine doors into a 12-in. start box, a 24-in. runway, and a 12-in. goal box. The alley had aluminum side and end walls and hardware cloth ceiling and floor. A metal food cup was permanently attached to the end wall of the goal box. The retaining cage was 9 in. in diameter and 4 in. high and was constructed of hardware cloth.

Interruption of a photocell beam located in the runway section of the alley, 6 in. beyond the start door, could, according to experimental plan, present the CS in both the alley and retaining cage. This arrangement allowed presentation of the CS to an *S* in the alley, following initiation of the running response, but prior to exposure to an empty food cup. At the same time an *S* in the retaining cage could receive the same duration of exposure to the CS. The CS was automatically terminated when an *S* was removed from the goal box of the alley.

Breaking the before-mentioned photocell beam also started a Standard Electric clock, which was stopped by the interruption of a second beam located in the goal box, 6 in. from the end wall.

Testing.—The hurdle box was made of aluminum with the exception of a grid floor and a clear plastic side wall. A drop door separated the hurdle box into two identical compartments, 7 in. high, 6½ in. wide, and 7 in. long. In its open position the door provided a 1-in. hurdle between the two compartments. Opening the door turned on the CS and started a Standard Electric clock. Both were terminated when *S* crossed the hurdle and activated microswitches supporting the grid floor.

The stabilimeter consisted of a modified, 1,000-gm., dietary scale, upon which was attached a small cage, 5 in. high, 5 in. long, and 3 in. wide. Three walls of the cage were made of clear plastic, the fourth wall and ceiling of hardware cloth, and the floor of stainless steel rods. Vertical displacements of the scale produced by *S*'s movements were linearly amplified and graphically recorded by means of a Hunter eyelid movement recording system. Positioned 2 in. to the rear of the cage, facing the hardware cloth wall, was an 8-in. loud speaker. The speaker was capable of delivering a brief, loud, 1,000-cycle tone, which had been found to consistently produce a measurable startle in naive *S*s. Associated programing equipment allowed presentation of the CS prior to the startle-producing tone.

---

[2] Stabilimeter *S*s were trained and tested by Neva L. Cummings and hurdle-box *S*s by Merle F. Goldstein.

Conditioned stimulus.—The CS consisted of interrupted noise and flashing light. Speakers located on the ceiling and floor of the refrigerator, above and below the experimental situations, provided diffuse white noise. Three-watt light bulbs arranged above, or in the ceiling of the several situations produced approximately equal illumination in each. This involved three bulbs above the runway, at points 6, 18, and 30 in. from the food cup, single bulbs above the center of the retaining cage and the stabilimeter, and a single bulb in the ceiling of each compartment of the hurdle box. The noise and light were presented in a train of $\frac{1}{2}$-sec. bursts separated by $\frac{1}{4}$ sec.

Procedure

Pretraining.—Seven days prior to experimental training *S*s were placed on a 24-hr. food deprivation schedule which was maintained throughout the experiment. Each *S* received 12 gm. Purina lab chow checkers in its individual home cage, 1 hr. after daily training, and free access to water. During the course of these 7 days *S*s were habituated to the experimental situation by systematic handling, exposure to reward pellets (.045 gm. each, purchased from the P. J. Noyes Company) scattered on a raised platform, exposure to the runway in groups of three, and finally two direct placements in the goal box of the alley with five reward pellets in the food cup.

Training.—On training days a yoked pair of *S*s was taken from its home cage in an adjoining room and brought to the experimental room. One *S* was placed in the retaining cage while the other *S* received a series of massed training trials in the alley. An alley trial was initiated by placing *S* in the start box. Approximately 3 sec. later the start door was raised allowing *S* to traverse the runway. When *S* entered the goal box, the goal-box door was lowered to prevent retracing. The *S* was removed from the goal box immediately after eating its five reward pellets, or in the case of nonreward trials, approximately 2 sec. after approaching the empty food cup.

Following the completion of each trial, *S* was placed in an open cage with water available for approximately 20 sec., while response time was recorded and the apparatus reset. The next trial was then run in the same manner. Each *S* received 4, 32, 32, 32, and 16 such trials on 5 successive days.

After the first *S* of each pair had completed its daily trials it was transferred to the retaining cage, and its yoked partner was given the same number of trials in the alley. For half of the pairs the experimental *S*, and for half the control *S*, was run first in the alley. All *S*s received the same 50% reinforcement schedule in the alley, + − + −, on Trials 1–4, and + − + − − + + − − − + + + − − + on each subsequent 16-trial block. The particular 16-trial sequence was selected by virtue of its being an irregular sequence which, when repeated, provides counterbalanced first-order sequential probabilities, i.e., nonreinforcement (or reinforcement) equally as often follows a reinforcement as it does a nonreinforcement.

Treatment of experimental and control *S*s differed only in that the CS was presented on each *nonreinforced* alley trial for experimental *S*s, but was never presented in the alley for control *S*s. Since the CS was simultaneously delivered to the retaining cage and alley, control but not experimental *S*s were exposed to the CS in the retaining cage.

Testing.—Following each pair of *S*s' last trial in the alley, one member of the pair was immediately tested in either the hurdle box or the stabilimeter. The other member was returned to its home cage until testing was completed for its mate, at which time it was tested in the same situation. For half the pairs the experimental *S*, and for half the control *S*, was tested first.

An *S* to be tested in the hurdle box was placed in one compartment with the door closed between the compartments. Two minutes later the door dropped, presenting the CS and allowing *S* to cross to the opposite side. When *S* crossed the hurdle the CS terminated and the door was raised. Ninety seconds later the door was again dropped presenting the CS and allowing *S* to return to its original side, again terminating the CS. Sixteen successive trials were run in this shuttle fashion, on each of which response time was recorded.

An *S* to be tested in the stabilimeter was placed in the stabilimeter cage and allowed 10 min. habituation before the first trial was run. The CS was then presented for 3,200 msec., the last 200 msec. of which the startle tone occurred and *S*'s displacement of the scale was measured. Twenty successive such startle trials were administered, separated by intertrial intervals of 2, 3, and 4 min. in a fixed randomized order.

Results and Discussion

Training.—The training procedures for experimental *S*s were designed to produce frustrative-nonreward, in accordance with its definition, and to allow presentation of the CS in a temporal relationship to it which would make the conditioning of frustration likely, if such occurs. Since these procedures involved systematic presentation of the CS during the latter portions of a response chain, on nonreinforced, but not on reinforced trials, they would also be expected to produce discriminative behavior over that section of the response chain. The appearance of a slowing in experimental *S* runway speeds on nonreinforced CS trials was, in fact, viewed as a prerequisite to testing, not only as assurance that the CS was a discriminable cue, but also as a necessary observation for the assumption that frustration had become conditioned to the CS. The latter qualification is based on the further assumption of frustration theory (Amsel, 1958) that conditioned frustration, when aroused by cues systematically paired with nonreward, tends to mediate learned or unlearned avoidance responses and is one of the factors inherent in discrimination learning involving frustrative-nonreward.

Figure 1 presents mean runway speeds for all experimental and control *S*s, plotted separately over successive blocks of reinforced and nonreinforced trials. There was little apparent difference between reinforced and nonreinforced trial speeds over the course of training for the control group, or over the first half of training for the experimental group. By the end of training, however, the approach response of the experimental group was appreciably slower on nonreinforced

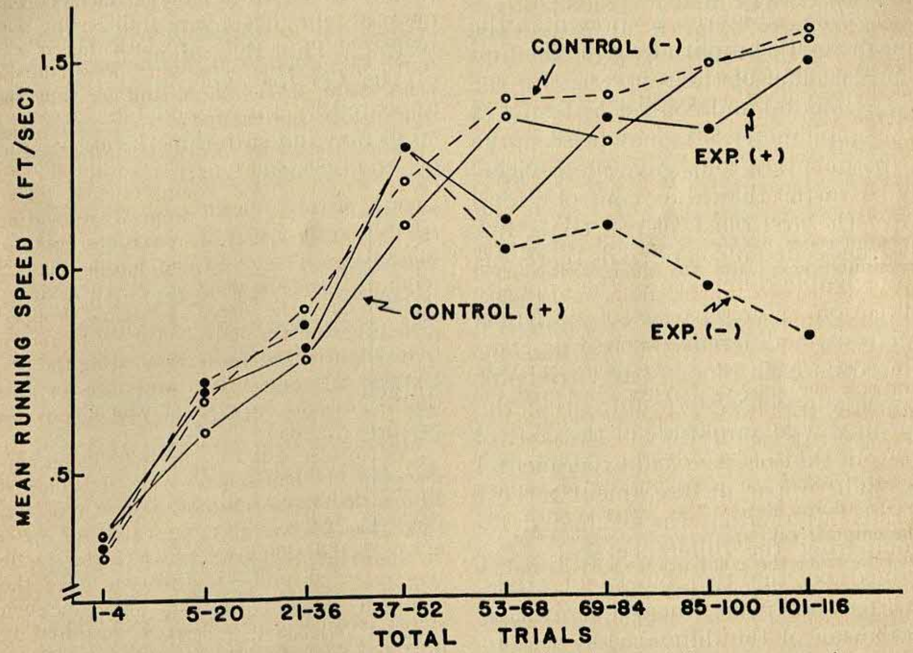


Fig. 1. Mean runway speeds for all experimental and control *S*s over successive blocks of reinforced and nonreinforced training trials.

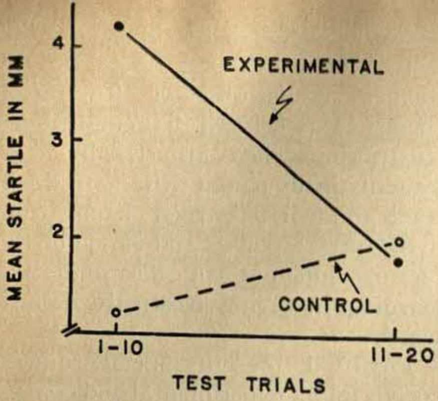


FIG. 2. Mean startle amplitudes over the first and last half of the test series for those experimental and control *S*s tested in the stabilimeter.

CS trials than on reinforced trials. The representativeness of the latter observation is witnessed by the fact that each of the 21 experimental *S*s obtained slower mean speeds over its nonreinforced than over its reinforced trials, on each of the last two 16-trial blocks of training.

Testing.—Figure 2 presents the mean startle amplitudes over the first and last half of the test series for the experimental and control *S*s tested in the stabilimeter. As may be seen, the experimental group gave more vigorous startle than the control group over the first block of test trials, but by the second block of trials there was a negligible difference in the opposite direction. For purposes of statistical analysis, a difference score was computed on each block of test trials for each pair of *S*s, by subtracting the mean startle amplitude of the control *S* from that of its yoked experimental *S*. The apparent Treatment × Test Trials interaction was evaluated by comparing the difference scores obtained on the two blocks of trials. A t test for related measures revealed the mean of the difference scores obtained over Trials 1–10 to be reliably different from the mean of the difference scores obtained over Trials 11–20, $t\ (8) = 2.68$, $p < .05$. Subsequent t tests showed the mean of the difference scores to be reliably different from zero over Trials 1–10, $t\ (8) = 2.73$, $p < .05$, but not over Trials 11–20, $t\ (8) = .27$.

The obtained difference in startle amplitudes between experimental and control *S*s over Trials 1–10 is in obvious agreement with the frustration theory prediction. The more vigorous reaction of experimental *S*s to the startle tone is consistent with the view that the CS produced a relative increase in drive level for those *S*s for which it had previously been paired with frustrative-nonreward. Not predictable was the degree of persistence of this potentiation. Exposure of experimental *S*s to the CS on test trials, without the primary event of frustrative-nonreward, represented extinction, and should have progressively diminished the effectiveness of the CS over the course of testing. That the potentiating effects of the CS for experimental *S*s disappeared by the second block of test trials may indicate that the extinction of conditioned frustration is char-

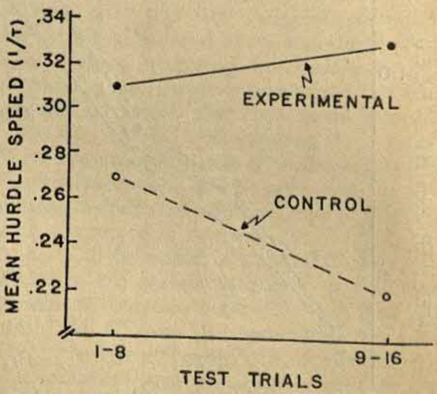


FIG. 3. Mean hurdle crossing speeds over the first and last half of the test series for those experimental and control *S*s tested in the hurdle box.

acteristically rapid, or merely that only a low level of conditioning had been attained by the start of testing.

Figure 3 presents the mean hurdle crossing speeds over the first and last half of the test series for experimental and control *S*s tested in the hurdle box. As may be seen, the experimental group responded faster than the control group over both blocks of trials, but this difference was more pronounced in the last half than in the first half of the test series. The latter observation is crucial to the proposal that CS cessation was uniquely reinforcing for experimental *S*s. The appearance of faster overall speeds for experimental *S*s could be attributed to the energizing effects of the CS previously demonstrated with the startle procedure. Differential reinforcement for hurdle crossing, on the other hand, implies a divergence of the hurdle crossing speeds of the two groups with additional hurdle crossing trials.

Statistical evaluation of the hurdle crossing speeds was carried out in a manner identical to that employed with the startle data. A difference score was computed on each block of test trials for each pair of *S*s, by subtracting the mean hurdle speed of the control *S* from that of its yoked experimental *S*. The apparent Treatment × Test Trials interaction was evaluated by comparing the difference scores obtained on the two blocks of trials. A t test for related measures revealed the mean of the difference scores obtained over Trials 9–16 to be reliably greater than the mean of the difference scores obtained over Trials 1–8, $t\ (11) = 2.47, p < .05$. Subsequent t tests showed the mean of the difference scores to be reliably greater than zero over Trials 9–16, $t\ (11) = 1.92, p < .05$, one-tailed, but not over Trials 1–8, $t\ (11) = .66$.

It is apparent from Fig. 3 that the reliable divergence over trials of the two groups is largely a result of a decrease in control group speeds. The small increase observed in experimental group speeds during testing did not approach statistical reliability. If it is assumed that other variables inherent in the test situation, such as a decrease in exploratory behavior, were operating to decrease *both* experimental and control hurdle crossing over test trials, then the observation that escape from the CS served to *maintain* the speed of hurdle crossing for experimental, but not for control *S*s, seems adequate to demonstrate its effectiveness as a reinforcer for experimental *S*s. At the same time, the failure to find an absolute increase in experimental *S* speeds indicates that this reinforcing effect cannot be considered large in comparison to such other variables.

Indeed the transitory nature of the startle findings would have made the observation of a large reinforcement effect surprising. Although, the many differences in the two testing procedures dictate some degree of caution in extrapolating from one to the other concerning the exact course of CS effectiveness over test trials, the startle data would at least suggest that the amount of reinforcement occasioned by CS cessation would have been minimal by the end of hurdle-box testing.

Regardless of the absolute magnitude or persistence of the effects, the obtained differences between experimental and control *S*s in both startle and hurdle crossing are in agreement with frustration theory predictions. When considered in conjunction, the results from the two testing situations appear to place the secondary drive status of conditioned frustration on somewhat firmer ground. To the extent that the present findings are similar to those obtained in investigations of fear (Brown & Jacobs, 1949; Brown et al., 1951; Miller, 1948) they provide further indication that nonreinforcement in appetitional reward situations may be difficult to distinguish in certain of its behavioral effects from other "aversive" stimulation.

REFERENCES

Amsel, A. The role of frustrative nonreward in non-continuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

Barry, H., Wagner, A. R., & Miller, N. E. Effects of alcohol and amobarbital on performance inhibited by experimental extinction. *J. comp. physiol. Psychol.*, 1962, **55**, 464–468.

Brown, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.

Brown, J. S., Kalish, H. I., & Farber, I. E. Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *J. exp. Psychol.*, 1951, **41**, 317–328.

Goodrich, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.

Spence, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Wagner, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

Wagner, A. R. Sodium amytal and partially reinforced runway performance. *J. exp. Psychol.*, 1963, **65**, 474–477.

(Received August 30, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 149–154

IMMEDIATE MEMORY FOR CONSONANTS AS A FUNCTION OF FREQUENCY OF OCCURRENCE AND FREQUENCY OF APPEARANCE[1]

JAMES H. KORN AND RICHARD H. LINDLEY

Carnegie Institute of Technology *Trinity University*

This study provided a test of the effect of 2 kinds of frequency on immediate memory for consonants. It was found that the number of consonants correctly recalled increased significantly with an increase in the frequency with which the consonant occurred in the English language (frequency of occurrence) and also with an increase in the frequency with which it appeared on any one trial during the experiment (frequency of appearance). The interaction was not significant. Errors from within a stimulus item occurred much more often than extraitem errors. The "spew hypothesis" of Underwood and Schulz was used to explain the differences in the frequency data. Analysis of the error data also provided further evidence for the existence of a response selector mechanism.

This study is concerned with the ect of two types of frequency on the mediate recall of consonants. The t type of frequency, which will be ed to as frequency of occurrence s the frequency with which a particular verbal unit has been experienced in the past history of an individual. Frequency of appearance (FA) is the other variable and is defined as the number of times a verbal unit appears among a string of such units presented to *S* for recall.

Underwood and Schulz (1960) have summarized the previous work on frequency, much of which has shown this variable to have a facilitating effect on learning and recall. In their own experiments, these authors found that frequency predicted quite well the learning of single-letter responses and responses of low integration. Most of this research, however, has been concerned with frequency of occurrence and not with frequency of appearance as defined above. Also, these variables have not been manipulated previously in an immediate recall situation. The purpose of the present investigation is to study the combined effects of these two types of frequency.

METHOD

Subjects.—The *S*s were 90 undergraduate students who served in the experiment as part of a course requirement. When *S*s reported for the experiment they were assigned to one of the six experimental groups according to a table which had been previously constructed in a random manner.

Materials.—Using the Underwood count of frequency of use of individual letters (Underwood & Schulz, 1960), the nine consonants of highest frequency (T, N, S, R, H, L, D, C, F) and the nine of lowest frequency (G, P, B, V, K, X, J, Z, Q) were selected. The letter M was not used because of its ready confusion with N, and W was not used since it is the only letter which takes more than one syllable to pronounce. Vowels were not used in order to avoid sequences which might form meaningful words.

All nine consonants were used in preparing the low FA items (e.g., FDLSHCTNR). To form the items, the consonants were arranged in alphabetical order and were assigned numbers from 1 to 9. A table of random numbers was then entered and the consonants were recorded as their respective numbers appeared in the

---

[1] Based upon a Master of Science thesis done at Carnegie Institute of Technology under the direction of the second author.

The authors would like to thank R. L. Ernst for his constructive criticism of this paper.

table with the restriction that each consonant be used only once in any one item. For the medium FA list, five consonants per item were used (e.g., LSTHDHSLT). Eighteen combinations of five consonants were selected in such a manner that each consonant would appear equally often in the final list of items. Each of the five consonants in each combination was then assigned a number from 1 to 5. Again, a table of random numbers was entered and consonants were recorded as their number appeared, this time with the restriction that each of four of the consonants be used twice in an item and the fifth consonant be used only once. The same method was used to prepare the high FA list in which only three different consonants per item were used (e.g., SSHCHSHCC): combinations of three consonants were selected, were assigned numbers from 1 to 3, and were arranged in items as their numbers came up in a random number table. Again, each consonant was used equally often in the list of items. After the items were prepared, they were typed individually in capital letters on plain white 3 × 5 in. cards.

Design.—The use of Underwood's frequency count and the preparation of items containing three, five, and nine consonants provided a 2 × 3 experimental design. Each *S* served in only one of these six conditions. Since each *S* was presented with each of the 18 stimuli in his condition, the 18 presentations can be conceived of as 18 trials and the result is a mixed design (Lindquist, 1953, Type III).

Apparatus.—A 1½ × 4 in. hole was cut in the center of a 23 × 20½ in. piece of ⅛-in. board. A sliding door arrangement was constructed on the front (*S*'s) side of the board. This door could be raised to expose the hole when *E* pulled a string which was attached to the top of the door and ran over a pulley fastened to the top of the board. All of the front portions of the apparatus were painted flat black. On the back of the board, a frame was constructed around the right and bottom sides of the hole. This frame was used to hold the 3 × 5 in. cards containing the stimulus items in such a manner that, when the door in front was raised, *S* would be presented with a string of nine consonants.

Procedure.—When *S* entered the room he was seated in front of a table containing the apparatus described above. In front of *S* were two sharpened pencils and a pile of 18 2¼ × 8½ in. strips of paper each of which contained nine short lines on which *S* would record his responses. The *E* took a seat behind the board and read the following instructions to *S*:

> This is a test of memory. A string of nine consonants will be exposed when the door on the panel in front of you is lifted. I will present the nine consonants and then lower the door. You are then to write the consonants on your answer sheet in the same order in which they were presented. Notice that your answer sheet consists of a number of slips of paper, each containing nine short lines, one line for each consonant. Consider the following string as an example, but don't write anything for this one. . . . (*E* presented the example). . . . You would then write this item on your answer sheet and would try to write the letters in the order in which they appeared. You will only have ten seconds in which to write the item so do not pause too long. If you can remember some of the consonants but not others, put a dash in the place of the ones you can't remember. When I say "read ," turn the slip of paper over and be prepar for the next item. Remember not to wr anything until the door is lowered. Do y have any questions?

If *S* had no questions, *E* inserted card in the frame on the back of paratus, raised the door for 6 sec., then lowered the door. The *S* was given 10 sec. to record his response, during which time *E* changed the cards in preparation for the next stimulus presentation. A stop watch was used to keep the time intervals fairly constant. After presentation of all 18 items, *S* was asked not to discuss it with anyone in his class. Between *S*s, *E* shuffled the stimulus cards so that each *S* received the items in a different order.

RESULTS

The mean number of correct consonants and *SD*s for the six groups are presented in Table 1. All differences are in the expected direction: The high FO groups recalled more consonants than the low FO groups and, as FA increased, number of consonants recalled increased.

An analysis of variance was carried out on the correct consonants. Both FO and FA are statistically significant; $F\ (1, 84) = 21.91$, $F\ (2, 84) = 25.96$, $p < .001$. The nonsignificant F for trials (the 18 successive presentations of the consonant items) is consistant with the hypothesis that there were no practice effects over the 18 trials.

The a priori scoring formula (Guilford, 1954, p. 448), $S = R - (W/K - 1)$, was applied to the data for each *S* to correct for the possibility of guessing. In this formula, S is the corrected score, R is the number of correct consonants, K is the number of different consonants in the item (three, five, or nine), and W is the number of committed errors (as opposed to omitted errors or blank spaces). The mean number of correct consonants after the correction for guessing and the *SD*s are given in Table 2. The analysis of variance of this corrected data again resulted in significant *F*s for both FO and FA; $F(1, 84) = 18.34$, $F(2, 84) = 6.94$, $p < .01$. However, the significant FA effect appears to be due almost entirely to the difference between high and medium FA conditions.

TABLE 1

MEANS AND *SD*s FOR CORRECT CONSONANTS OVER 18 TRIALS

FA	High FO		Low FO	
	M	*SD*	*M*	*SD*
High	114.80	15.77	106.33	13.95
Med.	97.00	16.82	78.93	11.80
Low	93.27	19.79	72.87	11.87

TABLE 2

MEANS AND *SD*s FOR CORRECT CONSONANTS CORRECTED FOR GUESSING

FA	High FO		Low FO	
	M	*SD*	*M*	*SD*
High	97.57	20.92	88.37	18.63
Med.	87.77	17.30	69.47	12.58
Low	88.03	20.48	67.05	11.47

The *S*s in the experiment made two kinds of errors: errors of commission (wrong letter) and errors of omission (blank spaces). The mean number of each kind of error and their *SD*s are given in Table 3. The differences between the groups for errors of commission are very small while the differences in errors of omission are large and are in the same direction as the means for correct responses. When an analysis of variance was performed on the errors of omission, both FA and FO were again shown to be statistically significant; $F(1, 84) = 16.10$, $F(2, 84) = 21.85$, $p < .001$. It will be recalled that *S*s had been instructed to leave blank spaces when they were uncertain about a particular response. This instruction was included in an attempt to restrict guessing. If this instruction was effective, the differences in group performance should be attributable largely to errors of omission, and it appears from the results given in Table 3 that this is indeed so.

Although the errors of commission do not correlate with performance,

TABLE 3

MEANS AND *SD*s OF ERRORS OF OMISSION AND COMMISSION

FA	High FO		Low FO	
	M	*SD*	*M*	*SD*
	Omission			
High	12.60	9.95	19.20	11.32
Med.	29.33	14.48	45.13	19.04
Low	31.47	18.70	51.06	16.60
	Commission			
High	34.20	13.11	36.73	10.27
Med.	35.06	10.79	37.33	16.72
Low	36.53	14.37	37.87	10.86

TABLE 4

FREQUENCY OF OCCURRENCE OF THE FOUR TYPES OF ERRORS OF COMMISSION

Type of Error	High FA		Med. FA		Low FA	
	H-FO	L-FO	H-FO	L-FO	H-FO	L-FO
Within-item intrusions	499	512	463	440	—[a]	—[a]
Other consonants in same list	13	39	58	108	—[a]	—[a]
Consonants from other lists	0	0	4	9	5	8
Vowels	1	0	1	3	2	0

[a] The first two types of errors are indistinguishable for the low FA groups. The total errors of these two types were 541 for high FO and 560 for low FO.

they are interesting from another point of view. An analysis was made to determine the source of these errors; i.e., whether they came from within the item presented for recall or from some other source outside the item. The frequency of each type of error is listed in Table 4. This analysis demonstrated that the vast majority of the errors were due to within-item intrusions. A second point to be noted is that extraitem intrusions occur much more often among consonants of low FO than among those of high FO.

As a further analysis of errors of commission, a count was taken of the number of times each consonant occurred as an intrusion for every other consonant. No explanation is readily available for the results of the count which is presented in Tables 5 and 6. There was no relation between FO and intrusions. Underwood and Schulz (1960) have determined how often each letter of the alphabet occurs as a response to every other letter, but this measure also bears no relation to these results except in one or two cases (e.g., K as an intrusion for J). In some cases similarity of pronunciation of the consonants would appear to be an accurate predictor (e.g., D and T, P and B). An interpretation in terms of phonetic elements is also possible in some instances. However, since the present experiment was not designed to yield a well-controlled error analysis, no further attempt will be made to explain the occurrence of intrusions.

TABLE 5

NUMBER OF TIMES EACH CONSONANT OCCURRED AS AN INCORRECT RESPONSE TO EVERY OTHER CONSONANT OF THE SAME FREQUENCY RANGE

Error	High Frequency Correct Response								
	T	N	S	R	H	L	D	C	F
T	—	14	7	6	18	24	36	30	21
N	23	—	26	45	29	17	22	18	42
S	12	21	—	16	26	35	20	27	33
R	12	20	19	—	15	31	12	8	29
H	15	35	27	20	—	16	10	21	17
L	15	26	12	30	23	—	14	24	14
D	41	25	26	8	22	10	—	15	17
C	34	19	29	10	20	31	15	—	11
F	28	37	26	14	24	7	23	13	—

TABLE 6

Number of Times Each Consonant Occurred as an Incorrect Response to Every Other Consonant of the Same Frequency Range

Error	Low Frequency Correct Response								
	G	P	B	V	K	X	J	Z	Q
G	—	9	18	13	9	16	7	25	23
P	17	—	28	22	24	12	22	35	28
B	35	38	—	27	20	21	17	25	10
V	34	25	32	—	38	35	23	22	17
K	15	24	21	17	—	19	41	22	14
X	22	21	21	29	27	—	19	21	20
J	29	18	16	30	37	19	—	14	20
Z	26	23	17	13	24	19	12	—	7
Q	38	27	11	14	15	14	23	12	—

Discussion

According to the "spew hypothesis" proposed by Underwood and Schulz (1960), "the order of availability of verbal units is directly related to the frequency with which the units have been experienced." Thus, the more frequently an individual has experienced a particular verbal unit, i.e., the greater the FO, the more readily will this unit become a response in a new situation. Extending this notion to the situation of the immediate recall of consonants leads to the hypothesis that consonants of high FO would be recalled better than consonants of low FO.

The results of this study have confirmed this hypothesis and have extended the Underwood and Schulz notions concerning frequency to an immediate memory situation and to a second concept of frequency. The position taken by these authors that frequency is important because it controls the availability of responses in recall has received additional empirical support. The idea that consonants which have been experienced more frequently are more readily available, is consistent with the greater number of consonants correctly recalled by the high frequency groups, and with the error data. These statements apply to both types of frequency with no significant interaction between them. The fact that there were fewer extraitem intrusions in the high frequency groups indicates that for these *S*s, the correct responses were more available, even though they were often misplaced within an item. To put it another way, since intrusions from outside an item occurred more often in the low frequency groups, the strength of the correct responses for these groups was not as great relative to the incorrect responses and the probability of some other consonant being emitted as the response was higher.

The preponderance of within-item intrusions in the error data also provides evidence for a selector mechanism in immediate recall. This refers to a response tendency or set produced by the instructions which causes *S* to respond with the stimuli from the items presented to him for recall. That is, incorrect responses, as well as the correct ones, are usually from the set of responses which *S* has been instructed to use and not from the other two possible sources: consonants in other experimental items and extraexperimental letters. This is true even though these other sources may contain consonants more frequently used in the daily verbal behavior of the individual.

However, although *S*s in this experiment limited their responses largely to the consonants from the item presented for recall, they did not do so to as great an extent as did *S*s in paired-associate studies reported elsewhere. In the present study, extraitem intrusions constituted 10% of all errors, while in paired-associate studies this value has been around 5%. This result was unexpected, since, "the maximum effect of the selector mechanism will occur when measurement is made immediately after stimulation [Underwood & Schulz, 1960, p. 305]." However, this decreased efficiency of the selector mechanism is probably due to the peculiar characteristics of the immediate memory situation. The rapidity of presentation of the items may have caused some confusion among the consonants since *S*s were sometimes still in the process of recording their responses when the next item was presented. It is also possible that certain letters or letter sequences served as stimuli to elicit consonant responses from other items. At any rate, it is apparent that the selector mechanism is not as selective in the immediate recall of consonants as in paired-associate learning.

REFERENCES

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received August 31, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 155–162

RETINAL AND ASSUMED SIZE CUES AS DETERMINANTS OF SIZE AND DISTANCE PERCEPTION [1]

J. C. BAIRD

Princeton University

3 experiments were conducted to determine the importance of the retinal image in perception of size and distance. Os were given 2 different types of instructions. One set of instructions (objective) emphasized judging physical size while the other set (analytic) required size judgments in terms of the retinal image. When other visual cues were eliminated, it was found that the assumed size of an object was an important determiner of its estimated size and distance under objective instructions but not under analytic instructions. When objective instructions were given, size and distance estimates were positively correlated but when analytic instructions were given no significant relationship was found between judgments of size and distance.

Holway and Boring in a classic study in 1941 investigated the perception of size with different available distance cues and concluded that accurate perception of an object's size is contingent upon accurate perception of its distance. In their study, as distance cues were successively eliminated, size estimates came to rely more upon retinal image size until finally, in the absence of distance cues, judgments depended exclusively upon the retinal image subtended by the object. These findings have been the source of at least two theoretical controversies and a host of experimental inquiries.

First, Holway and Boring's statement that size must depend upon the retinal image when distance cues are eliminated has been opposed. Ittelson (1951, 1953) has cogently argued that if *O* knows from past experience an object's physical size this knowledge coupled with a given retinal image size will permit accurate estimation of the object's distance. The issue, however, is not settled concerning the effect of familiar size upon apparent distance, and a number of writers assert that familiar size is not a cue to distance if *relative size* (change or difference in the size of objects of similar shape) is eliminated (Epstein, 1961; Gogel, Hartman, & Harker, 1957; Hochberg & Hochberg, 1952, 1953; Hochberg & McAlister, 1955). It still has proven difficult to explain results of several studies by recourse to a relative size hypothesis. For instance, in experiments by Hastorf (1950) and Smith (1952) relative size was held *constant* (in terms of the aforementioned definition) and yet the apparent physical distance of an object was shown to depend to a certain extent upon its assumed size; which to paraphrase Hastorf (1950), is the distal physical size believed to be associated with a particular retinal image pattern. If it were established that known or assumed size was a distance cue, theories of depth perception based solely upon the operation of innate perceptual mechanisms (Hochberg & Hochberg, 1952) would no longer be feasible.

A second aspect of the Holway and

---

[1] Based on a thesis submitted to the faculty of the University of Delaware in partial fulfillment of the requirements for the MA degree. The author wishes to thank F. L. Smith, W. L. Gulick, G. A. Cicala, and John T. Lanzetta for their helpful suggestions, and Charles E. Marvil for his assistance in conducting Exp. I.

Boring (1941) study challenged by recent *E*s has been the ability of *O*s to make retinal image matches. More specifically, investigators have asked if retinal image size supplies information for the determination of the size of an object whose physical size and distance is unknown. Several studies in addition to that of Holway and Boring have supported the thesis that size judgments under reduced distance cues approach an image match of the comparison and the standard stimuli (e.g., Chalmers, 1952; Hastorf & Way, 1952; Lichten & Lurie, 1950; Over, 1960). Nevertheless, Wallach and McKenna (1960) have interpreted their data as showing that the retinal image of an unknown object is not a size cue. These writers suggest that image matches obtained in earlier studies may have resulted from *O*s' assumptions that the standard stimulus was located at the same distance as the comparison stimulus. Wallach and McKenna also postulate that the initial size of a comparison stimulus presented *with* distance cues determines size judgments of a standard stimulus presented *without* distance cues. If Wallach and McKenna's notions were verified, it could be argued that separation of size and distance perception had not been experimentally demonstrated, a conclusion of wide theoretical consequence (Epstein, Park, & Casey, 1961; Gibson, 1952; Gilinsky, 1955).

It appeared to the present author that discussion concerning the original Holway and Boring study stemmed from one basic question: Is it possible in visual perception to recognize the relative areas of the visual field occupied by objects? An affirmative answer to this question is necessary if one is to claim that known or assumed size is a cue to distance; or that, size is determinate in the absence of distance cues. Hence, three experiments were undertaken to clarify the effects of retinal image and assumed size cues on size and distance perception. In view of recent experimentation (Carlson, 1960, 1962; Gilinsky, 1955; Jenkin & Hyman, 1959) different instructional sets were used, *objective* and *analytic*, and *O*s were required to estimate the size and distance of stimuli presented under minimal distance cues. For different *O*s the same stimuli were given different assumed size meanings while other *O*s received no information regarding the physical dimensions of the stimulus whose size and distance they were estimating.

EXPERIMENT I

Purpose.—Experiment I was designed to study the effect of analytic and objective instructions upon size and distance estimates of an object when *O* was led to assume its physical size. Three hypotheses were tested:

1. Under minimal distance cues, except for retinal image size, if objective instructions are given, size judgments will be a function of the assumed size of an object.
2. Under minimal distance cues, if analytic instructions are given, size judgments of an object of a given assumed size will be a function of the retinal image size subtended by that object.
3. Under minimal distance cues, except for retinal image size, if either objective or analytic instructions are given, distance estimates of an object of a given assumed size will correlate positively with the size judgments of that object.

Method

Observers.—Forty-eight undergraduate men served as *O*s, and each was naive concerning the purpose of the experiment. Each *O* had uncorrected normal vision as tested by an orthorater.

Apparatus and stimuli.—In a darkened room canvas was hung from the ceiling to enclose a well lighted rectangular area, 14½ × 3½ ft. At one end of the enclosure a combination headrest and chin rest oriented *O*'s head for viewing a comparison stimulus located inside the enclosure and 10 ft. from the headrest. The comparison was an equilateral triangle (30-in. altitude) constructed of white posterboard with a brightness of 5.8 ft-L as measured with the Macbeth illuminometer. When *O* placed his head in the first headrest, he could vary the size of the comparison by operating a balanced weight system. The canvas behind the triangle was black while the remainder of the enclosure was white.

A second headrest at the same height as the first changed *O*'s line of sight 90° to the right toward the standard stimulus which was on a table 25 ft. from the eye of *O* when he looked monocularly through a reduction tube. The standard stimulus source was obtained by removing the front side of a cardboard box and replacing it with white paper. Two 7½-w. incandescent bulbs located inside the box served to illuminate the front surface with a brightness of .46 ± .03 ft-L as measured with the Macbeth illuminometer. When viewed from 25 ft., the surface appeared textureless. Black cardboard sheets with the area of the standard figures cut out were hung singly in front of this illuminated surface in a plane perpendicular to *O*'s line of sight. The standard stimuli consisted of three rectangles and one triangle of light. One rectangle was the size of an ordinary foot ruler (i.e., 12 × 1 in.). The second rectangle was one half the size of a foot ruler, and the third rectangle was twice as large as a foot ruler. The triangle was equilateral with an altitude of 12 in.

Procedure.—Each of the 48 *O*s was randomly assigned to one of eight conditions. Six *O*s were given objective instructions and 6 were given analytic instructions in each of three ruler groups (small, normal, large) and in one no-ruler group which served as a control. The *O* viewed all stimuli monocularly (right eye). Each of the *O*s in the ruler groups was shown one of the standard rectangles of light and was told it was the size of a 12-in. ruler. The rectangle was then removed and the standard triangle was presented at the same distance. The *E* told *O* that the triangle was at the same distance as was the previous rectangle, and then *O* judged the size of this triangle by adjusting the comparison stimulus. In the no-ruler group *O*s judged the size of the triangle without being informed of its assumed size; that is, without first being shown the ruler. Instructions for the ruler groups were modified accordingly for these *O*s. In judging size *O*s were not allowed to look back and forth between the two figures, and they were asked to come to a decision as quickly as possible. The *O*s given objective instructions were directed in part:

> . . . adjust the size of the comparison by raising one of these weights until the two triangles are physically equal. By physically equal I mean if you were to take a ruler and measure both triangles they would be exactly the same measured size. . . .

The analytic instructions were similar to what Carlson (1960, 1962) has termed "apparent size" instructions. The *O*s given analytic instructions were directed in part:

> . . . adjust the size of the comparison by raising one of these weights until the two triangles appear equal in visual size. I am not interested in the measured physical sizes of the triangles, these may or may not be equal. . . . Moreover, while making your judgment, do not take distance factors into consideration. . . .

A method of adjustment was used, and each *O* made two size judgments, one ascending (initial size of comparison ½ in.) and one descending (initial size of comparison 29½ in.). One half of the *O*s in each condition was run on descending trials first and one half on ascending trials first. When each *O* completed his size judgment, he estimated (verbally) the distance of the standard triangle and then that of the comparison triangle, which was set at a constant height of 6 in.

Results and Discussion

Size judgments.—If the *O*s in the ruler conditions accurately judged the relative image sizes of the ruler and standard triangle, the assumed size of the triangle in the normal ruler group should have been 12 in.; in the large ruler group, 6 in.; and in the small ruler group, 24 in. The mean size judgments obtained for the *O*s given objective instructions were: normal ruler, 9.3 in.; large ruler, 6.3 in.; small ruler, 18.2 in.; and no ruler, 7.5 in. A retinal image match of standard and comparison triangles would result in a size judgment of

$4\frac{4}{5}$ in. across all groups. The mean size judgments for the *O*s given analytic instructions were: normal ruler, 8.2 in.; large ruler, 7.3 in.; small ruler, 8.6 in.; and no ruler, 7.8 in.

The data were subjected to an analysis of variance. The main effects of treatment groups and instructions were significant sources of variance ($p = .005$) but order of presentation was not. Direction of comparison stimulus movement (ascending or descending) also was significant ($p = .001$) but the only significant interaction was between treatments and instructions ($p = .001$). Further differences among means were tested by Duncan's new multiple-range test (Edwards, 1960). Under objective instructions the mean judgments for the larger ruler, normal ruler, and no-ruler groups each differed significantly from the mean judgment of the small ruler group ($p = .001$), and the normal ruler and large ruler group means differed significantly ($p = .05$). The only significant difference found between analytic and objective means within an individual treatment group was in that of the small ruler ($p = .001$). Other comparisons were not significant.

The hypotheses concerning size estimation were generally supported by the data. Size judgments by *O*s under analytic instructions were not significantly influenced by the assumed size of the standard while objective group judgments were. On the other hand, the differences between instructions were not marked within all groups, and in the analytic and no-ruler conditions overestimates were obtained. These latter findings may be attributed in part to the operation of the comparison stimulus, the height of which *O* had to decrease 24.5 in. on descending trials and increase only 4.3 in. on ascending trials to attain retinal image matches.

Distance estimates.—The mean distance estimates obtained for *O*s given objective instructions were: normal ruler, 19.7 ft.; large ruler, 9.1 ft.; small ruler, 23.3 ft.; and no ruler, 10.1 ft. Mean distance estimates for analytic instructions were: normal ruler, 12.0 ft.; large ruler, 13.5 ft.; small ruler, 11.6 ft.; and no ruler, 10.3 ft.

Using combined data from all four groups correlation coefficients were determined between size (mean of ascending and descending judgments) and distance estimates independently for analytic and objective instructions. In all cases the null hypothesis tested was that the population correlation was zero. Under objective instructions a correlation of .47 ($df = 23$, $p = .05$, two-tailed test) was obtained between size and distance estimates of the standard, while under analytic instructions the correlation between size and distance was .19 which is not significant. The correlation between distance estimates of the comparison and size judgments of the standard was $-.04$ for objective instructions and $-.005$ for analytic instructions. Correlations between distance estimates of the standard and comparison stimuli revealed an $r = .53$ ($df = 23$, $p = .01$, two-tailed test) with analytic instructions and an $r = .04$ with objective instructions. Distance estimates of the comparison stimulus were all very close to the actual distance of the comparison and hence will not be reported.

The hypothesis concerning correlations of size and distance was not entirely borne out: under objective instructions size and distance estimates were positively correlated but this was not the case for *O*s given analytic instructions. These findings corroborate data reported by Carlson (1960) and are understandable from at least two theoretical positions. First, it may be that the assumed size of the standard triangle *was not* a sufficient cue to distance, and that *O*s judged distance on the basis of relative size differences between the standard and comparison stimuli. In the present case of objective instructions such differences, after *O*'s size judgments, would lead to a positive correlation of size and distance. In the analytic groups, because size judgments were almost identical, distance estimates tended to be highly similar with individual differences obscuring any significant size-distance relationships.

However, an alternative explanation of the present data might state that *O*s *were using* the assumed size of the standard triangle to estimate distance under objective instructions but not under analytic instructions. The reason *O*s under analytic set did not use the assumed size cue would be that they were previously directed to disregard distance factors in coming to a decision of size and were in a no-cue situation concerning distance.

EXPERIMENT II

The distance estimates of Exp. I can be accommodated equally well by either the relative size or known size hypothesis. Experiment II was undertaken to test the known size hypothesis by finding out if the apparent physical distance of a visual stimulus can be shown to depend upon its assumed physical size.

Method

Observers.—Thirty undergraduate men served as *O*s, and each was naive concerning the purpose of the experiment. The visual selection criteria were the same as in Exp. I.

Apparatus.—The apparatus was the same used in Exp. I. However, a canvas screen was hung in front of the headrest which faced the comparison triangle, thus occluding it from *O*'s view when he entered the enclosure. The stimuli were the three standard rectangles used in Exp. I. The size of the comparison was set at a constant height of 6 in. and the distance of all stimuli were the same as in Exp. I.

Procedure.—The 30 *O*s were randomly separated into three equal groups. Each *O* estimated the physical distance of one of the rectangular strips of light, whose measured size he was told was the same as a foot ruler, 12 in. high and 1 in. wide. As in Exp. I, *O*s viewed the standard monocularly through a reduction tube. Ten *O*s verbally estimated the distance of the large, 10 of the normal, and 10 of the small ruler. After this estimate, *E* removed the screen in front of the headrest which faced the comparison stimulus, and *O* estimated its distance.

Results and Discussion

All standard stimuli were presented at a distance of 25 ft. Thus, prediction from the known size-apparent distance hypothesis (Epstein et al., 1961) is that the apparent distance of the normal ruler would be 25 ft.; of the large ruler, 12.5 ft.; and of the small ruler, 50 ft. The mean distance estimates obtained for the three groups were: normal ruler, 24.0 ft.; large ruler, 14.3 ft.; small ruler, 48.2 ft. However, the standard deviations obtained were: normal, 7.3 ft.; large, 5.6 ft.; and small, 42.6 ft. The mean estimates of the comparison triangle were: normal, 13.1; large, 13.2; and small, 11.6 with standard deviations of 3.0, 4.1, and 2.6, respectively.

A Kruskal-Wallis (Senders, 1958) test revealed significant differences between the estimates of the standard by the three groups ($p = .025$). Rank-sums tests (Senders, 1958) were carried out for each pair of ruler groups. The null hypothesis that the groups' estimates had been drawn at random from populations with the same median was rejected in the case of large ruler vs. normal ruler ($p = .05$) and for large ruler vs. small ruler ($p = .01$) while the null hypothesis could not be rejected for the normal vs. small comparison.

The results of Exp. II indicate that the assumed size of a stimulus may determine its apparent physical distance though the effect is subject to wide *O* differences. These results support the thesis that known size (Ittelson, 1951, 1953) or assumed size (Hastorf, 1950; Smith, 1952) is a cue to distance and are contrary to conclusions drawn from studies in which it was found that such cues were not effectively used in distance perception (Epstein, 1961; Gogel et al., 1957; Hochberg & Hochberg, 1952; Hochberg & McAlister, 1955).

EXPERIMENT III

Because of the postulated importance of the comparison stimulus in earlier experiments (Wallach & McKenna, 1960, Exp. I), an investigation was conducted to determine the effect of a comparison stimulus presented with distance cues upon the size and distance estimates of a standard stimulus presented with minimal distance cues.

Method

Observers.—Sixteen undergraduate men served as *O*s, and each was naive concerning the purpose of the experiment. The visual selection criteria were the same as in Exp. I.

Apparatus.—The apparatus was the same as in Exp. I. The canvas screen was placed in front of the headrest which faced the comparison triangle. The standard was the 12-in. triangle which was located on a table 20 ft. from the eye of *O* when he looked monocularly through the reduction tube.

Procedure.—Each of the 16 *O*s was randomly assigned to one of two equal groups. Eight *O*s received analytic instructions and 8 received objective instructions. Four *O*s in each instruction group received ascending trials and 4 received descending trials, and each *O* made only one size judgment. The *O* was instructed in the operation of the adjustable weights while viewing their shadows through the curtain. On a given trial *O* viewed the standard for 5 sec., *E* removed the screen in front of the comparison, and *O* made his adjustment. The *O* had the *option* of leaving the comparison stimulus as he initially saw it, of making it larger (ascending group), or smaller (descending group) to match the standard. The *O* was not allowed to look back at the standard once the screen covering the comparison was removed. Hence, it is unlikely that *O*s were able to implicitly locate the standard and comparison stimuli at the same distance. The comparison was set at ½ in. for ascending trials and 12 in. for descending trials. After his size judgment, *O* estimated (verbally) the physical distance of the standard and comparison figures. One half the *O*s in each group estimated the distance of the comparison first and one half estimated the distance of the standard first.

Results and Discussion

If *O*s were matching retinal images, the size of the comparison triangle should have been set at 6 in. The mean size judgments obtained were 4.7 in. for objective instructions and 5.3 in. for analytic instructions. An analysis of variance revealed no significant differences between these values; though, as in Exp. I, size judgments varied as a function of direction of the comparison movement ($p = .005$).

The mean distance estimate of the standard was 6.2 ft. for objective instructions and 8.9 ft. for analytic instructions while distance estimates of the comparison stimulus closely approximated its actual distance.

Correlation coefficients were determined between size and distance estimates independently for analytic and objective instructions. Under objective instructions a correlation of .74 ($df = 7$, $p = .05$, two-tailed test) was obtained between size and distance estimates of the standard, while under analytic instructions the correlation between size and distance was $-.21$ which is not significant. The correlation between distance estimates of the comparison and size judgments of the standard was $-.72$ ($df = 7$, $p = .05$, two-tailed test) for objective instructions and .12 for analytic instructions. Correlations between distance estimates of the standard and comparison stimuli revealed an $r = .65$ ($df = 7$, $p = .10$, two-tailed test) with analytic instructions and an $r = -.47$ with objective instructions.

As in Exp. I, the initial setting of the comparison stimulus was found to be a strong determiner of size judgments. These results corroborate those of Wallach and McKenna (1960). However, size judgments of the standard did *approach* a retinal image match of standard and comparison despite the fact that *O*s could not locate the two stimuli at the same distance, and despite the fact that *O*s were allowed the option of not moving the comparison from its initial position.

The finding that size judgments under reduced cue conditions do not vary as a function of instructions is consonant with an earlier study by Over (1960) while the result that size-distance correlations are related to instructions agrees with data from Exp. I,[2] and with Carlson's (1960) study. Several reasons for the correlations could be advanced. The most feasible explanation of the distance estimates obtained in the present study is that they were dependent upon the size settings of the comparison. That is, the relative image sizes of the standard and comparison stimuli were not constant for all *O*s, and an *O*'s initial setting of the comparison was perceived as too small (ascending trials) or too large (descending trials) when he was given the opportunity to look at the standard a second time. These relative size differences became clearly discernible in *O*s' distance estimates of the standard. Consequently, the reason *O*s did not perfectly match retinal images (standard and comparison) was due to the methodology involved and not to the impossibility of the task. On the other hand, *O*s given analytic instructions were either not "set" to perceive the relevance of the relative size cue or had no reason to believe the two triangles were the same physical size (see instructions to analytic group).

General Discussion

Several authors have been reluctant to admit that the size of the retinal image is an important factor in visual perception (Gibson, 1952; Hochberg & Hochberg, 1952; Wallach & McKenna, 1960). Contrary to this position, data from the present study indicate that retinal image size can be effectively used for the estimation of both size and distance. These findings lend support to a wide range of previous experimentation most of which has been discussed earlier in this paper.

It seems that much disputation concerning the importance of the retinal image in perception has its source in semantics. A case in point is the theoretical separation of relative and familiar size. The distinction loses much of its force if one considers that in *any* situation where estimates of either size or distance are obtained a comparison stimulus of some sort is being used (Boring, 1952). Viewed in this way, differences between experiments purporting to demonstrate the sole effects of relative size (e.g., Exp. III of the present study), and experiments presumably demonstrating the sole effects of familiar or assumed size (Gogel et al., 1957; Exp. II of the present study) become differences between the length of time intervening between presentation of the comparison and standard stimuli. Once it is recognized that a relative size cue is employed even *when the size or distance of a single stimulus is being judged,*[3] the important question becomes whether this information alone will permit accurate estimation of the relative size or distance of objects. A number of carefully designed studies have answered this question affirmatively (Epstein, 1961; Gogel et al., 1957; Hochberg & Hochberg, 1952; Hochberg & McAlister, 1955). Nonetheless, data obtained in the present study with analytic instructions suggest that the relative size cue is not always effective in determining either the apparent physical size (Exp. I) or apparent physical distance (Exp. III) of a visual stimulus. A similar statement could be proposed for the cue of assumed size.

---

[2] One reversal of the correlations found in Exp. I was obtained. In Exp. III under objective instructions size estimates were negatively correlated with distance estimates of the comparison stimulus ($p = .05$). The reason for this reversal is not clear though Jenkin and Hyman (1959) have reported a similar correlation.

[3] This statement is partially based upon the supposition that recognition of the size equality of two stimuli is recognition of their relative sizes.

Results of the present series of experiments also are germane to the recent theoretical issue regarding the validity of the size-distance invariance hypothesis (Epstein et al., 1961). The present study lends support to this hypothesis, at least for objective instructions. Although the invariance hypothesis has undergone criticism from several quarters (Epstein, 1961; Epstein et al., 1961; Gruber, 1954; Jenkin & Hyman, 1959), it is this writer's view that its employment should be reserved for situations in which objective instructions are used. Moreover, the statement of invariance concerns *one* visual stimulus, not two stimuli subtending different size retinal areas (Jenkin & Hyman, 1959). Perhaps we would do well to delimit the application of the invariance hypothesis rather than persist in attempts to encompass all size-distance perception under its auspices.

REFERENCES

BORING, E. G. The Gibsonian visual field. *Psychol. Rev.*, 1952, **59**, 246–247.

CARLSON, V. R. Overestimation in size-constancy judgments. *Amer. J. Psychol.*, 1960, **73**, 199–213.

CARLSON, V. R. Size-constancy judgments and perceptual compromise. *J. exp. Psychol.*, 1962, **63**, 68–73.

CHALMERS, E. L. Monocular and binocular cues in the perception of size and distance. *Amer. J. Psychol.*, 1952, **65**, 415–423.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

EPSTEIN, W. The known-size-apparent-distance-hypothesis. *Amer. J. Psychol.*, 1961, **74**, 333–346.

EPSTEIN, W., PARK, J., & CASEY, A. The current status of the size-distance hypotheses. *Psychol. Bull.*, 1961, **58**, 491–514.

GIBSON, J. J. The visual field and the visual world: A reply to Professor Boring. *Psychol. Rev.*, 1952, **59**, 149–151.

GILINSKY, A. S. The effect of attitude upon the perception of size. *Amer. J. Psychol.*, 1955, **68**, 173–192.

GOGEL, W. C., HARTMAN, B. O., & HARKER, G. S. The retinal size of a familiar object as a determiner of apparent distance. *Psychol. Monogr.*, 1957, **71** (13, Whole No. 442).

GRUBER, H. E. The relation of perceived size to perceived distance. *Amer. J. Psychol.*, 1954, **67**, 411–426.

HASTORF, A. H. The influence of suggestion on the relation between stimulus size and perceived distance. *J. Psychol.*, 1950, **29**, 195–217.

HASTORF, A. H., & WAY, K. S. Apparent size with and without distance cues. *J. gen. Psychol.*, 1952, **47**, 181–188.

HOCHBERG, C. B., & HOCHBERG, J. E. Familiar size and the perception of depth. *J. Psychol.*, 1952, **34**, 107–114.

HOCHBERG, C. B., & HOCHBERG, J. E. Familiar size and subception in perceived depth. *J. Psychol.*, 1953, **36**, 341–345.

HOCHBERG, J. E., & MCALISTER, E. Relative size vs. familiar size in the perception of represented depth. *Amer. J. Psychol.*, 1955, **68**, 294–296.

HOLWAY, A. H., & BORING, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21–37.

ITTELSON, W. H. Size as a cue to distance. *Amer. J. Psychol.*, 1951, **64**, 188–202.

ITTELSON, W. H. Familiar size and the perception of depth. *J. Psychol.*, 1953, **35**, 235–240.

JENKIN, N., & HYMAN, R. Attitude and distance-estimation as variables in size-matching. *Amer. J. Psychol.*, 1959, **72**, 68–77.

LICHTEN, W., & LURIE, S. A new technique for the study of perceived size. *Amer. J. Psychol.*, 1950, **63**, 281–282.

OVER, R. The effect of instructions on size-judgments under reduction conditions. *Amer. J. Psychol.*, 1960, **73**, 599–602.

SENDERS, V. L. *Measurement and statistics.* New York: Oxford Press, 1958.

SMITH, W. M. Past experience and the perception of visual size. *Amer. J. Psychol.*, 1952, **62**, 389–403.

WALLACH, H., & MCKENNA, V. V. On size-perception in the absence of cues for distance. *Amer. J. Psychol.*, 1960, **73**, 458–460.

(Received September 4, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 163–167

REACTION TIME TO SINGLE AND TO FIRST SIGNALS[1]

ROBERT GOTTSDANKER, LEE BROADBENT, AND CLARKE VAN SANT

University of California, Santa Barbara

Choice reaction times were measured for 6 adult Ss by 2 procedures. There was a Single-Choice condition in which, after a fixed warning interval, S was required to move a lever away from himself or toward himself according to which of 2 signal lamps was lighted. In the double-choice condition, there followed ½ sec. after the 1st signal a 2nd choice involving 2 other signal lamps and a choice by the other hand. For each S, mean reaction time for the Single-Choice condition was reliably shorter than that for the 1st choice in the Double-Choice condition. This was regarded as corroboration of Poulton's thesis that a person's manner of response, including latency, is influenced by his expectations of the requirements of the immediate future.

It has been emphasized by Poulton (1950) that any explanation of the interaction of responses to successive signals solely in terms of how an earlier response affects a later one, e.g., the *psychological refractory period*, is inadequate in its neglect of anticipation. He states: "Present performance is always dominated by the subject's idea of the immediate future, by what he expects to happen, and by what he is trying to accomplish [p. 112]." In accounting for the initial reaction time of a complex response, he notes: "The whole performance is dominated by the anticipatory mental set or aim of the subject, and to this extent is a single whole [p. 104]." Three experimental findings were used to substantiate this position. First, reaction time for the start of a pattern-tracing task was about 25 msec. longer than that for the start of a simple unidirectional movement. Second, when an *unexpected* "stop" signal was given at a corner point in a tracing path, time to stop was 250 msec. longer than "ordinary complex graded reaction time." Third, when an *unexpected* "continue" signal was given at a point where *S* expected to stop, his normal reaction time was increased by 350 msec. in starting the continuation.

Only the first of these findings provides anything like clear-cut evidence for Poulton's thesis, even though it imputes the smallest effect to anticipation. In the other two cases, the apparent increases in reaction time could be attributed to the time required to overcome muscular force patterns existing at the instant of the unexpected signal. Curiously, Poulton employed still another procedure which might have been expected to demonstrate a difference attributable to anticipation, but which did not. In this, *S* moved a key in one direction to the first auditory signal and in the opposite direction to the second. Fixed time intervals of from 100 to 800 msec. were used for blocks of trials. Comparison of reaction time to the first signal with that to a single signal revealed no significant difference. Here, apparently, the initial response was *not* influenced by the expected nature of the whole task.

This last finding is in disagreement with incidental observations made in

---

[1] Acknowledgment is given to the support of this study by the National Science Foundation. L. Broadbent and C. Van Sant were participants in a program of undergraduate research participation, Grant NSF-15649, under the supervision of R. Gottsdanker.

the Santa Barbara Laboratory for a somewhat similar task. It was noted that when *S* was required to make two successive choice responses, even though fairly well separated in time, he appeared to have a reaction time for the first choice which was longer than normal. The present experiment represents a refinement of this observation. Comparison was made between reaction time for a single choice and reaction time for the same choice when *S* knew that it would be followed by a second choice.

It is important in the successive-choice procedure that an appropriate time interval be selected. If the interval between signals is too long, the two responses become completely separate tasks; i.e., there must be a limit to the interval which will yield the anticipation effect, if any. If the interval is too short, the second signal per se might influence the first response. This effect has been reported by Vince (1948) to occur occasionally and has been shown to be of real importance by Ellson and Hill (1948). The interval selected here was $\frac{1}{2}$ sec. With this interval, the response is completed before the onset of the second signal, so a direct effect is impossible. Fortunately, it proved to be short enough for *S*'s anticipation to influence the first response.

METHOD

Apparatus

Display and controls.—On a flat table before *S* were mounted four neon bulbs in a 1-in. square. Midway between the two lamps on the left, and 1 in. farther to the left was a short Plexiglas lever. It was to be moved with thumb and index finger in the direction of whichever of the two left-hand lamps went on, away from *S* (upward) or toward him (downward). A similar lever was placed to

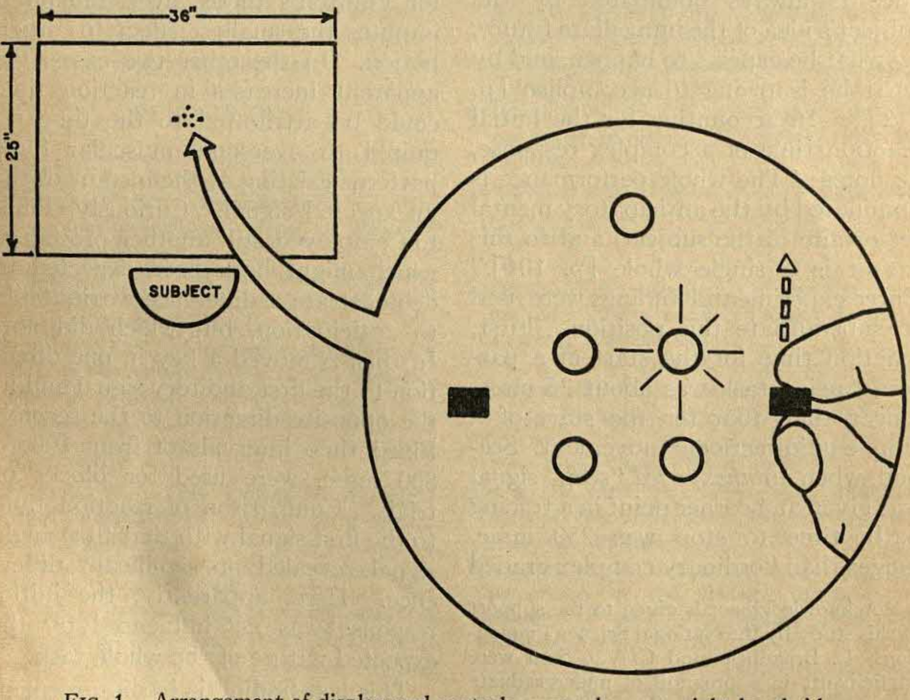


FIG. 1. Arrangement of display and controls; upper lamp on right-hand side shown flashing and correct motion of response lever indicated.

the right of the right-hand pair of lamps. There also was a fifth neon bulb which was used as a warning light, at the horizontal center of the display and slightly over an inch above the far (or top) lamps. The *S* was seated in a sound treated and air conditioned room. The arrangement is shown in Fig. 1.

Programing devices.—Time between trials, interval between warning light and first signal, and certain auxiliary operations were controlled by means of a set of motor-driven cams, each of which activated a microswitch. The ½ sec. between the first and second signal was generated by an electronic time-delay relay. A flash of approximately 20 msec. for each signal was produced by direct current through a series capacitor.

Recording devices.—Presentation of a signal simultaneously started the operation of an electronic counter. When *S* moved the appropriate lever a short distance in the correct direction, it stopped operation of the counter. One electronic chronoscope was used to indicate the reaction time to the first signal and another for the reaction time to the second signal. Times were recorded to the nearest millisecond.

Procedure

Conditions.—There were two conditions tested, Single-Choice and Double-Choice.

Single-Choice condition. One second after the warning lamp flashed, one of the two lamps, of either the left-hand or right-hand pair flashed. The *S* was told in the instructions, before a block of trials, which pair would be used. The only unknown factor was which of the two lamps would flash. His task was to move the response lever as quickly as possible in the direction away or toward him, depending on which lamp flashed.

Double-Choice condition. The *S* was instructed that on each trial there would be two choices, one on the left-hand side and one on the right-hand side. He was told which of the two sides would be first. One second after the warning lamp flashed, one of the two lamps on the first side flashed; ½ sec. later one of the two lights on the other side flashed. The *S*'s task was to move each appropriate lever as quickly as possible in the correct direction. The Double-Choice condition, then, was exactly the same as the Single-Choice condition with the exception of the second signal and response.

Conduct of the experiment.—Trials were run in homogeneous blocks of the four subconditions: Single-Choice, left-hand; Single-Choice, right-hand; Double-Choice, left-hand first; Double-Choice, right-hand first. Ordinarily, there was one session a day composed of all of the 40 trials that were given of a subcondition, the series of sessions taking place on consecutive days. Throughout, the interval between trials was 15 sec., and the daily session required about 15 min.

Before the start of the experiment proper (the four test sessions), there were two practice sessions of similar length, in which all subconditions were equally represented. Also, before each test session there were eight "warm-up" trials. Serial effects were controlled by appropriate counterbalanced and random orders. If *S* made a mistake on a trial, that trial was placed at the end of the sequence and was repeated so as to give a total of 40 good trials. Practically identical overall percentages of error were found for the Single-Choice and Double-Choice conditions, the respective values being 5.2% and 5.6%. No information was given to *S* about his performance.

Subjects.—Six undergraduate volunteers from the University of California, Santa Barbara, were tested. There were four men and two women, all between 19 and 23 yr. of age. They were tested individually, usually in the early evening for 6 consecutive days.

RESULTS

The variable of concern in this experiment was condition: Single-Choice or Double-Choice. The measure of concern was first reaction time (in the Double-Choice condition) or single reaction time (in the Single-Choice condition). Table 1 presents the mean single and first reaction times in milliseconds for each *S*, under each subcondition, with an additional breakdown for direction of correct choice.[2] Of particular importance is the comparison of each entry in the

---

[2] Reaction times to the second signal are incomplete. It was found during the course of the study that the measurement was inaccurate. However, a sufficient amount of accurate data was obtained subsequently to ascertain that the times were typical of such studies, about 25 msec. shorter than those for the first response (Davis, 1956; Elithorn & Lawrence, 1955). No further use was intended for this measure.

TABLE 1

MEAN REACTION TIMES (in msec.) FOR SINGLE AND FOR FIRST CHOICES

S	Single-Choice				Double-Choice			
	Left Side		Right Side		Left Side		Right Side	
	Up	Down	Up	Down	Up	Down	Up	Down
1	289	256	298	245	309	272	309	278
2	251	232	218	203	256	262	214	262
3	218	235	207	251	312	322	270	255
4	275	238	262	223	302	262	299	270
5	217	254	211	204	254	286	229	227
6	304	309	324	285	409	378	437	336
M	259	254	253	235	307	297	293	271

Note.—$N = 20$ for each entry.

Single-Choice condition with the corresponding entry in the Double-Choice condition (e.g., S_1 had a mean reaction time in the Single-Choice condition for left-hand upward response of 289 msec. as compared with a mean of 309 msec. for the same response when it was the first choice of the Double-Choice condition). Note that in every comparison save one, the mean reaction time for the Single-Choice condition is shorter than that for the first reaction time under the Double-Choice condition. The only exception occurred with S_2, right side, upward direction. The mean reaction time under the Single-Choice condition was 4 msec. longer than that for the same response under the Double-Choice condition (218 msec. vs. 214 msec.).

TABLE 2

RESULTS OF *F* TESTS BETWEEN SINGLE AND FIRST REACTION TIMES FOR EACH *S*

S	Error Variance (Within Groups, $df = 152$)	F (1, 152)
1	3,146	5.11*
2	2,375	8.65**
3	1,981	78.34***
4	2,557	17.73***
5	3,141	9.74**
6	7,579	37.73***

*$p < .05$.
**$p < .01$.
***$p < .001$.

Table 2 shows F tests from the $2 \times 2 \times 2$ triple classification analysis of variance of single and first reaction times for each *S*. The method described under "Case X" by McNemar (1955, p. 328) was used. As is seen, all *S*s showed a significant effect for the condition variable.

DISCUSSION

It was demonstrated beyond question that reaction time to a choice is lengthened when *S* expects to make a subsequent choice which is signaled $\frac{1}{2}$ sec. after the first. This dependence of present response on expectation of the near future may be described, without any real interpretive extension, in terms of inclusiveness of the task: reaction time is longer when the response becomes the initial part of a more complex operation. This is exactly in line with the thesis proposed by Poulton.

Moreover, in the present study, unlike in Poulton's procedures, there was no possibility that the retarding of the initial response of the complex condition was because of past utilization or further requirements of precision by the response unit. Lengthening of reaction time occurred because *S* had a subsequent task to perform with a different response unit. The two movements became part of a unified complex solely because they were channeled through a single operator.

There remains the question of *why* the latency of a response is increased when it is made the first part of a more complex act. It could be said that *S* is not able to "concentrate" as well when he must make a second response also. This would put the explanation on the same footing as those of why *S* is not able to respond rapidly with an uncertain or unfavorable foreperiod or in choice situations, i.e., because of a lesser degree of "preparedness." Unfortunately, we do not yet appear to have adequate methods for making independent measurements of states of "concentration" or "preparedness" nor of "attention" or "activation." Of particular interest here is the suggestion that *S* selects the level of this state (whatever its nature) which is appropriate to the task which confronts him.

REFERENCES

DAVIS, R. The limits of the "psychological refractory period." *Quart. J. exp. Psychol.*, 1956, **8**, 24–38.

ELITHORN, A., & LAWRENCE, C. Central inhibition: Some refractory observations. *Quart. J. exp. Psychol.*, 1955, **7**, 116–127.

ELLSON, D. G., & HILL, H. The interaction of responses to step function stimuli: I. Opposed steps of constant amplitude. *USAF AMC memo. Rep.*, 1948 (Nov.), No. 694-2P.

MCNEMAR, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

POULTON, E. C. Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, 1950, **2**, 99–112.

VINCE, M. The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, 1948, **38**, 149–157.

(Received September 4, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 168–176

BEHAVIOR IN A CONTINUOUS-RESPONSE TASK WITH NONCONTINGENT REINFORCEMENT

SEYMOUR ROSENBERG[1]

Bell Telephone Laboratories, Murray Hill, New Jersey

The purpose of the study was to examine and test a number of fundamental properties of a linear stochastic model formulated for a continuous-response task. The continuous-response task used in this study is one in which responses and outcomes occur in ordered pairs over a sequence of discrete trials. 64 female Ss were run for 150 trials using a noncontingent schedule to determine trial outcomes. The fundamental assumption of linearity was explored with the data and was found to be tenable for most Ss in the sample. The presence of a 2nd fundamental property predicted by the model, termed heteroscedasticity, was tentatively supported by the data. Predictions concerning asymptotic response means, variances, and 3rd central moments were all verified. Parameters of the model were estimated.

A continuous-response task is one in which the set of available response alternatives and the set of outcome values are each ordered along a spatial or other clearly specified continuum. Of particular interest in this paper is the type of continuous-response task in which responses and outcomes occur in ordered pairs over a sequence of *discrete trials*. Empirical studies of individual learning in this type of task are exemplified by the line-drawing experiments and by a number of discrete-positioning studies in the motor-learning literature (Bilodeau & Bilodeau, 1961).

The models proposed by Suppes (1959, 1960), which are extensions of statistical learning theory, appear to be the first formal treatment of the continuous-response task with discrete trials. The present paper describes a linear stochastic model based on somewhat different notions of behavior change from those used in the linear and stimulus-sampling models described by Suppes. The basic notion of the model in the present paper is that *S* changes his response in proportion to the distance between the actual response and the correct response on a given trial. If the correct response changes from trial to trial *S* is, in effect, "tracking" successive positions of the correct response with a lag of one trial. A similar notion is present in a model recently proposed by Anderson (1961). Model I by Anderson and the one described here are both essentially feedback models. A notion incorporated in the present model is that the tracking rate may vary randomly from trial to trial instead of being fixed over trials. A description of the model will be deferred until the task, the procedure, and the experimental design have been described.

The main purpose of the paper is to test a number of basic properties of the model using two different noncontingent reinforcement schedules. A second purpose of the paper is to examine the generality of the model across two different sets of instructions to *S*s. In one set, *S*s were given standard instructions, viz., *S*s were told to guess the correct response on a trial. In the second set, *S*s were told that the correct response on a trial depended upon the simultaneous response of another *S*. Results obtained with the latter instructions are particularly relevant in extensions of the task and model to two-person interactions.

---

[1] The writer wishes to acknowledge the helpful suggestions and advice of R. Gnanadesikan and J. W. Tukey in the preparation of this paper. Thanks are also due to S. Cotten Wallen for assistance in data collection and processing and to Ray Kudlick for apparatus design.

METHOD

Subjects and apparatus.—Sixty-four female clerks and typists (modal age, 18 yr.) at Bell Telephone Laboratories were scheduled as *S*s according to the hour of the day that they could be spared from their normal work.

The apparatus consisted of two identically equipped booths. The walls of the booths were sufficiently high so that when 2 *S*s were run simultaneously they could not see each other. Each booth contained equipment sketched in Fig. 1. A response knob in the lower left corner of the tilted panel permitted *S* to select any number between 0 and 100. To select a number *S* turned the knob counterclockwise to a point on the scale. Different colored lights on the same panel served as signals for *S* to respond.

At the far end of the booth, a DuMont 401A oscilloscope displayed the trial outcome as a small illuminated dot on the face of the cathode-ray tube. The only part of the tube face visible to *S* was an 8-cm. square in which the dot appeared. The *S* controlled the horizontal displacement of the dot by the size of the number *S* selected and *E* controlled the vertical displacement of the dot. The bounds on *S*'s response continuum (0 and 100) corresponded to the extreme left and extreme right sides of the square, respectively. A single black line .04 cm. thick was painted

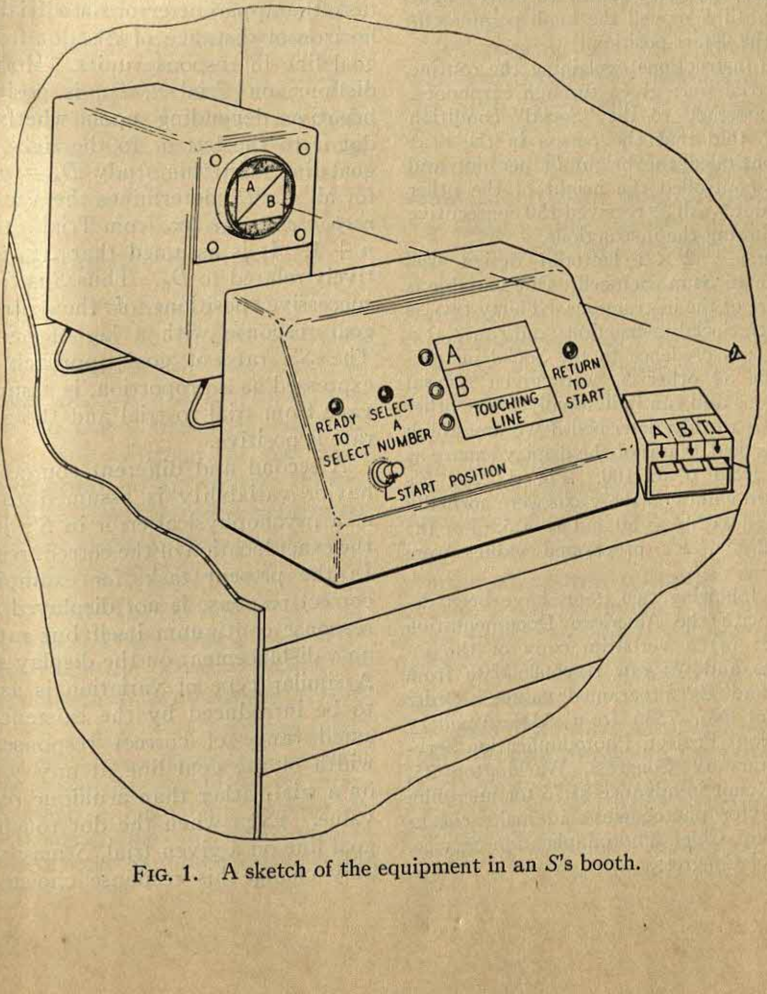


FIG. 1. A sketch of the equipment in an *S*'s booth.

diagonally across the square as shown in Fig. 1. A correct outcome was defined for *S* as the dot touching anywhere on the line, referred to as the goal line.

The *S* also reported the position of the dot on each trial by depressing one of three switches mounted in a box on the right side of the table. A corresponding light on the panel lit automatically to confirm to *S* that her report had been received by *E*. Switches A and B corresponded, respectively, to Subareas A and B on the display square. The switch labeled TL (spelled out "touching line" on the actual apparatus) was to be pushed by *S* when she saw the dot touch the goal line.

Procedure.—Each trial lasted at least 12 sec. and was partitioned as follows: (*a*) yellow "ready to select" light for 1 sec.; (*b*) green "select a number" light for 5 sec. or until *S* made a selection, whichever was later; (*c*) delay of 1 sec.; (*d*) display of the outcome dot for 3 sec. during which *S* reported the position of the dot; (*e*) red "return to start" light blinked until *S* moved the knob pointer off-scale to the "start position."

Taped instructions explaining the routine on each trial were given through earphones. The *S*s assigned to the "Social" condition were also told that the person in the next booth controlled the height of her dot and that she controlled the height of the other person's dot.[2] All *S*s received 150 consecutive trials following the instructions.

Design.—A 2 × 2 factorial design was used with 16 *S*s in each cell. One variable is the nature of the instructions. Thirty-two *S*s were given Social instructions and run 2 at a time to lend credence to the Social instructions; the 32 other *S*s were given Asocial instructions and run individually. The other variable is readily described if we first think of the vertical, *y* axis of the display square as ranging from 0 to 100. For 32 *S*s, *E*'s programed values on the *y* axis were normally distributed with $\bar{y} = 50$ and $\sigma_y = 5$. For the other 32 *S*s, *E*'s programed values were normally distributed with $\bar{y} = 50$ but $\sigma_y = 15$. Four random sequences of 150 Gaussian deviates each were obtained from the Rand Tables to program *y* values. A sequence appropriately transformed in mean and variance was randomly assigned to *S* with the restriction that all sequences occur equally often in each cell of the 2 × 2 design.

BASIC PROPERTIES OF THE MODEL

Basic properties of the model are summarized in this section. However, detailed mathematical derivations from the model will not be given in this paper because of space limitations. A discussion of the model as applied to actual two-person interactions is available elsewhere (Rosenberg, 1962).

It is assumed in the model that *S* responds primarily to the magnitude of the apparent error. In the oscilloscope task the apparent error on a trial is the horizontal distance of *S*'s dot from the goal line in response units. Horizontal distance on Trial *n*, D_n, is positive or negative depending upon whether the dot is to the left or to the right of the goal line. In this study $D_n = y_n - x_n$ for all *n*. D_n determines the size of the *response change* Δx_n from Trial *n* to Trial $n + 1$. It is assumed that Δx_n is positively related to D_n. Thus *S* is tracking successive positions of the correct or goal response with a lag of one trial. The *S*'s rate of goal approach, when expressed as a proportion, is assumed to vary from trial to trial and the average rate is positive.

A second and different source of behavior variability is assumed to occur as a psychophysical error in *S*'s judging the exact location of the correct response. In the present task, for example, the correct response is not displayed on the response continuum itself but rather as an *x* displacement on the display square. A similar type of variation is assumed to be introduced by the existence of a small range of correct responses (the width of the goal line at any *y* value) on a trial rather than a unique response value. Even when the dot touches the goal line on a given trial, *S* may reasonably change his response choice somewhat on the next trial because the width of dot and goal line are consistent with this latitude in behavior. The *S*s generally report the dot "touching line" when $y_n = x_n \pm 5$.

---

[2] The following two items have been deposited with the American Documentation Institute: (*a*) a verbatim copy of the instructions and (*b*) raw response data from all *S*s' and *E*s' programed values. Order Document No. 7536 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit in advance $1.75 for microfilm or $2.50 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

The following difference equation summarizes these descriptive notions.

$$\Delta x_n = x_{n+1} - x_n = \alpha_n(D_n + \delta_n) \quad [1]$$

Here α_n may be identified as the rate of goal approach and δ_n includes the various estimation errors in making response x_{n+1} as well as psychophysical errors associated with *S*'s observing the *n*th outcome. Note that D_n affects Δx_n linearly. The following properties are assumed for the random variables α_n and δ_n: At each trial the values of α_n and δ_n, which determine the response x_{n+1}, are drawn independently and randomly from distributions depending only upon *S* and not upon the responses or outcomes of Trials 1, 2 . . . *n* for that *S*. In addition to these mathematically convenient assumptions it will be helpful to make the psychologically reasonable assumption that δ_n has a symmetrical distribution with zero mean for each *S*.

RESULTS AND DISCUSSION

Linearity.—An analysis of variance design was used to explore linearity and curvilinearity in the data of individual *S*s. A $(D, \Delta x)$ pair was selected on every fifth trial starting with Trial 80.[3] Fifteen $(D, \Delta x)$ pairs were thus obtained for each *S*. Every fifth trial was used in order to increase independence among the observation pairs within an *S* (Suppes & Frankmann, 1961). The independent variable *D* was divided into intervals of 11 units each and corresponding Δx values classified accordingly. The variability in Δx was partitioned into between-classes (intervals in *D*) and within-classes variance. The Δx values were distributed among at least three classes of *D* intervals for each of the 64 *S*s so that the between-classes variance in each analysis could be decomposed into linear and curvilinear components.

The presence of a linear relationship was tested by obtaining an *F* ratio of linear *MS* to within-classes *MS*, MS_w. The presence of a linear relationship is clearly established with $p < .05$ for 33 out of the 64 *F*s. The average of the 64 linear coefficients was 0.359. The linear coefficient is closely related to the estimate of $\bar{\alpha}$. The latter is discussed in the next section.

An overall test for curvilinearity was performed by obtaining an *F* for each *S* of the curvilinear *MS*, MS_c, to MS_w. The 64 *F*s are summarized in Fig. 2. Each point represents the results from an individual *S*'s data. The heavy lines which partition the graph into subareas according to *p* level are not exact (but are adequate for partitioning the points) since *df*'s vary among the *F* ratios. A contingency χ^2 between σ_y conditions and three ranges of *p* ($p < .05$, $.05 < p < .50$, $p > .50$) was not significant ($p > .30$). The overall distribution of points in Fig. 2 does not suggest the presence of a large curvilinearity effect although there is an excessive number of points at the lower *p* levels. Because the data from the same *S* were used in each analysis of variance it is not certain that the points in Fig. 2 correspond exactly to an *F* distribution. For example, an overall χ^2 test (by a log transformation of the *p* levels) of the distribution of the 64 *F*s would be extremely sensitive to the three *F* ratios with $p < .001$ and

---

[3] The second half of the experiment was used to study linearity since any systematic change in $\bar{\alpha}$ over time could result in an apparent curvilinear relationship between Δx and *D*. It was assumed that any changes in $\bar{\alpha}$ would have occurred in the early part of the experiment.

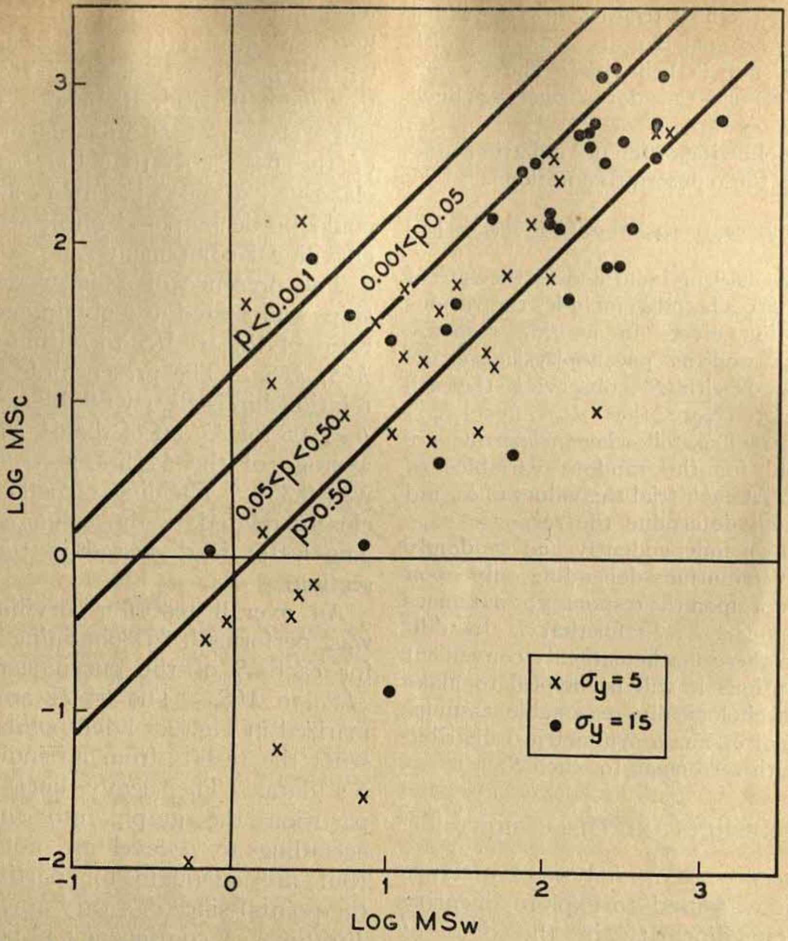


FIG. 2. A plot of the log *MS* attributable to curvilinearity (log MS_c) vs. the log *MS* within groups (log MS_w) obtained from the analyses of variance of an individual *S*'s data. (Diagonal lines correspond to critical values of *F*.)

would be somewhat misleading about the distribution of the remaining *F*s.

The curvilinearity question was explored further by calculating orthogonal quadratic coefficients for 64 *S*s and orthogonal cubic coefficients for the 45 *S*s whose 15 Δx values were distributed over at least four classes of *D* intervals (Wishart & Metakides, 1953). The average values of these two coefficients were both 0.000 with a range of −0.048 to 0.042 for the quadratic and of −0.003 to 0.003 for the cubic. Of the two coefficients the cubic is a more likely contributor to curvilinearity; however, the results indicate that this source of curvilinearity is very small.

Estimate of $\bar{\alpha}$.—A method for estimating $\bar{\alpha}$ for each *S* was derived from Equation 1. This estimate is essentially an unweighted least-squares estimate of the slope of a linear relationship betwee Δx and *D* with the assumption that the intercept is zero.

$$\hat{\bar{\alpha}} = \sum_n (\Delta x_n D_n) / \sum_n (D_n^2) \qquad [2]$$

Using Equation 2 $\hat{\bar{\alpha}}$ was calculated from the ungrouped data of each *S* for Trials 1–75 and 76–150. Results are summarized in Fig. 3. In all 128 estimates of $\bar{\alpha}$ one $\hat{\bar{\alpha}}$ was <0 and two $\hat{\bar{\alpha}}$'s were >1.00, specifically, $-.01 < \hat{\bar{\alpha}} < 1.09$.

Results of an analysis of variance on the $\hat{\bar{\alpha}}$ values are summarized in Table 1. Sources C or C × D were used as error terms for all *F*s. The assumption that $\bar{\alpha}$ is completely stationary is refuted (Source D). With a noncontingent schedule and with $D_n = y_n - x_n$, no rate of goal approach is consistently reinforced after $E(x_n) = \bar{y}$ and the decrement in $\hat{\bar{\alpha}}$ may indicate the extinction of the larger α_n values. In general, it would not be surprising to find that $\bar{\alpha}$ can be modified by specialized reinforcement schedules. The absence of a significant difference between σ_y conditions (Source σ_y) is important and in accord with the model. The significant difference between *S*s within conditions (Source C) indicates that $\bar{\alpha}$ is a consistent property of *S*. Finally, the larger $\hat{\bar{\alpha}}$ values from the Social condition are significantly different from $\hat{\bar{\alpha}}$ values from the Asocial condition (Source B).

TABLE 1

ANALYSIS OF $\hat{\bar{\alpha}}$ FOR EACH *S* IN EACH CONDITION FOR TRIALS 1–75 AND 76–150

Source	*df*	*F*
Individuals (A)	63	2.71***
σ_y	1	<1
Instructions (B)	1	4.77*
σ_y × B	1	<1
Within conditions (C)	60	2.62***
Trial blocks (D)	1	21.58***
A × D	63	
σ_y × D	1	2.03
B × D	1	1.27
σ_y × B × D	1	<1
C × D	60	

Note.—See also Fig. 3.
* $p < .05$.
*** $p < .001$.

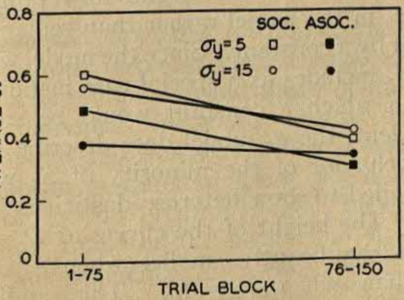


FIG. 3. Average $\hat{\bar{\alpha}}$ in each condition in each of two trial blocks.

Heteroscedasticity.—The conception of α_n as a random variable rather than as a constant implies that the dispersion of Δx values about a line (on an individual *S*'s Δx-D scattergram) with a slope equal to the $\bar{\alpha}$ of that *S*, is at a minimum when $D = 0$ and gradually increases as $|D|$ increases. However, if one were to use a straight line to obtain these residuals the obtained heteroscedasticity may be an artifact of any curvilinearity that is present in the Δx-D relationship. To preclude the possibility of this artifact, residuals were obtained from the *actual mean* of the Δx's for each *S* in each trial block (1–75 and 76–150) and in each of the *D* intervals. The root *MS*s of these residuals are plotted in Fig. 4. The extremes are generally based on less than five observations.

According to the model, parabolas are expected in Fig. 4 with a minimum at $D = 0$. The effect is most apparent when $\sigma_y = 5$ according to Fig. 4. The dropoff at the extremes of the curves when $\sigma_y = 15$ may be due to the bounds upon Δx imposed by the bounds of the response continuum. The plots in Fig. 4 are subject to a possible artifact from individual differences in response variability. That is, the residuals for the smaller *D* values are averaged over all *S*s while the aver-

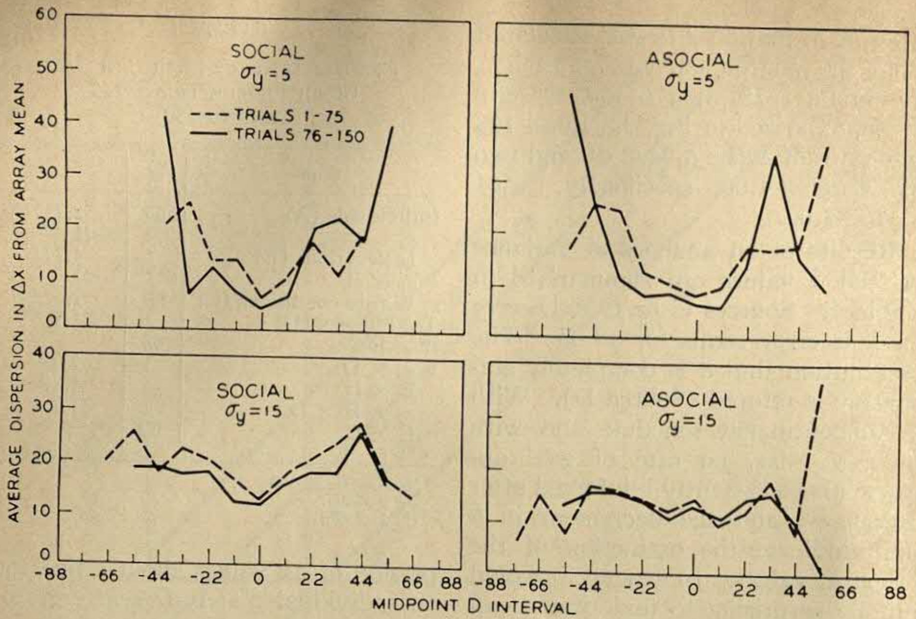


FIG. 4. Root *MS* of residuals in Δx from array means.

age residuals for larger *D* values represent primarily the more variable *S*s. This possible source of artifact was checked for Trials 76–150 by examining plots of high and low variability *S*s separately. Of the eight curves, two from each condition, six were definitely parabolic. Thus, it seems unlikely that individual differences in variability account entirely for the parabolic shape of most of the curves in Fig. 4.

An estimate of the pervasiveness of heteroscedasticity among individual *S*s was performed as follows. For each *S* the *MS* of the residuals in the interval, $D = 0 \pm 5$, was compared with the *MS* of residuals in all remaining *D* intervals together. Forty-seven out of 64 *S*s in Trials 1–75 and 38 *S*s in Trials 76–150 demonstrated smaller *MS* for $D \cong 0$ than for $|D| > 0$. Both relative frequencies are in the right direction.

A χ^2 test ($df = 1$) using a theoretical split of 32-32 was significant for Trials 1–75 only. However, the χ^2 test is somewhat conservative and approximate for two reasons. (*a*) The χ^2 test is two-tailed and the prediction from the model is, strictly speaking, one-tailed. (*b*) The ratio of the two *MS*s is approximately an *F* distribution and the probability that $F > 1$ is not, in general, .50 but depends on the *df*'s. Unfortunately, an exact test is not possible since the *df*'s vary among *S*s. The fact that the parabolas from Trials 76–150 seem genuine also argues for retaining α_n in the model rather than replacing it by a constant. Since the model with α_n includes as a special case instances in which a constant would be sufficient, the α_n model does describe the behavior of the minority of *S*s who failed to show heteroscedasticity.

The height of the curves at $D = 0$ is significantly smaller when $\sigma_y = 5$ than when $\sigma_y = 15$. This finding indicates that the size of one or more model parameters (not $\bar{\alpha}$, according to the previous section) depends on this experimental condition.

Asymptotic response moments.—Trials 76–150 were considered satisfactory for examining asymptotic behavior since no systematic change in the first four response moments was observed in this trial block. A number of predictions can be derived from the model without estimating parameters from the data. These predictions are examined in this section and can be briefly summarized as follows. With the assumption that δ_n is symmetrical, the asymptotic distribution of responses in all conditions is expected to have a mean the same as $\bar{y}$, a variance proportional to σ_y^2, and to be symmetrical in shape. These properties should be realized regardless of the shape of the distribution of α_n. Differences in $\bar{\alpha}$ due to instructions and trial blocks do not alter the predictions of asymptotic response means and skewness but may affect response variance and kurtosis.

To derive the moments of response, Equation 1 is first rewritten:

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n(y_n + \delta_n) \quad [3]$$

A difference equation for the rth raw moment is obtained by raising both sides of Equation 3 to the rth power and taking expected values. Necessary restrictions on α_n and δ_n for convergence of all moments are assumed.

Consider first the prediction that asymptotic response mean is equal to $\bar{y}$. The obtained grand mean of the 64 individual response means in Trials 76–150 was 49.9 which is not significantly different from $\bar{y}$ (49.3 in the actual program), the t being greater than the 0.1 level. No significant differences in means among conditions were found in a 2 × 2 analysis of variance. All 64 means fell within the range 41–59.

Figure 5 is a graph of the average response variance obtained in trial blocks of 10. Each point is an average of the individual *S*'s response variance in the trial block. Since no difference between Social and Asocial conditions was apparent, data from these two conditions were combined to provide more stable curves. A constant difference between σ_y conditions at asymptote is required by the model and is verified in these data.

The observed relationship between the asymptotic response variance and σ_y^2 would probably be expected even without this specific model unless one hypothesized that *S*s eventually optimize in the face of noncontingent reinforcement, i.e., come to choose $\bar{y}$ consistently. If *S*s had optimized in this manner the only response variance remaining at asymptote would be the variance of δ_n and the two curves in Fig. 5 would have converged to a common value.

The mean response variance in Trials 76–150 is significantly smaller than σ_y^2 at the .05 level according to a t test (the variances were first subjected to a log transformation). This finding is contrary to the prediction by Suppes and Frankmann (1961) if they intended their model to apply to the oscilloscope task. Thirty-nine (20 in $\sigma_y = 5$ condition) out of 64 *S*s have variances less than σ_y^2 in Trials

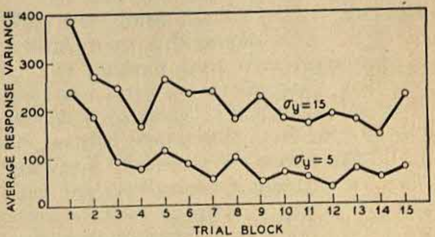


FIG. 5. Average response variance in blocks of 10 trials.

76–150. By contrast Suppes and Frankmann found 11 out of 60 *S*s with response variances less than that of the reinforcement distribution.

It is predicted from the model that the asymptotic distribution of responses is symmetrical if the distribution of y_n is symmetrical. According to a t test the mean of the 64 third central moments (divided by the 3/2 root of the variance) in Trials 76–150 is not significantly different from zero, the p level being greater than 0.1. Differences among conditions were also not significant.

An analysis of the fourth moment revealed that the typical distribution of asymptotic responses is significantly leptokurtic. The *S*'s asymptotic response distribution does not "match" the normally distributed y_n according to analyses of the response variance data and the kurtosis data. Matching is not required by the model for these two moments.

REFERENCES

ANDERSON, N. H. Two learning models for responses measured on a continuous scale. *Psychometrika*, 1961, **26**, 391–404.

BILODEAU, E. A., & BILODEAU, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–280.

ROSENBERG, S. Two-person interactions in a continuous-response task. In J. Criswell, H. Solomon, & P. Suppes (Eds.), *Mathematical methods in small group processes.* Stanford: Stanford Univer. Press, 1962. Pp. 282–304.

SUPPES, P. A linear model for a continuum of responses. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959.

SUPPES, P. Stimulus-sampling theory for a continuum of responses. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences, 1959.* Stanford Univer. Press, 1960.

SUPPES, P., & FRANKMANN, R. W. Test of stimulus sampling theory for a continuum of responses with unimodel noncontingent determinate reinforcement. *J. exp. Psychol.*, 1961, **61**, 122–132.

WISHART, J., & METAKIDES, T. Orthogonal polynomial fitting. *Biometrika*, 1953, **40**, 361–369.

(Received September 4, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 177–186

SUBJECTIVE SCALING OF LENGTH AND AREA AND THE MATCHING OF LENGTH TO LOUDNESS AND BRIGHTNESS[1]

S. S. STEVENS AND MIGUELINA GUIRAO[2]

Harvard University

With a device that allowed O to control the length of a line or the size of a square projected on a screen, magnitude and category scales were produced for the attributes, longness and largeness, and their inverses, shortness and smallness. The magnitude scales are power functions of the respective stimuli, and the inverse aspects are power functions with negative exponents (i.e., reciprocal functions). For length the exponent is about 1.0; for area about .7. The variable line was used to make cross-modality matches between apparent length and both loudness and brightness, as well as their inverses, softness and dimness. The results are similar to those obtained when the match is made with numbers (magnitude estimation).

Judgment of the apparent length of lines is one of the simplest and easiest judgments a person can make. These judgments are not undeserving of study, however, because it is sometimes possible with an easier task to discern features that are hard to detect in judgments of magnitude on more difficult continua. In addition there is the attractive possibility that apparent length may be used as the standard or criterion continuum against which to scale other continua. In practice we now scale most continua against number, by the method of magnitude estimation, but some critics have expressed uneasiness about numerical estimates. In the present experiments we therefore undertook to match apparent length to loudness and to brightness in order to see whether such cross-modality matching would prove useful. We also explored the relations between the category scale and the magnitude scale for length and area. These comparisons were carried out for both the magnitude and its inverse.

METHOD

A line 1 cm. wide and of variable length (or a square of variable size) was projected on a gray screen (183 cm. wide × 244 cm. high) placed 220 cm. in front of the observer (*O*) in a dimly lighted room. The length of the line (or the size of the square) could be controlled by *O* by means of a knob that operated a double rack and pinion to change a carefully machined aperture in a slide projector. The square was oriented with two opposite vertexes horizontal. The luminance of the gray screen was 65.0 db. re 10^{-10} L. The luminance of the line was 68.0 db. and of the square 66.3 db. These values provided a low but comfortable contrast between the line (or square) and the background.

Ten *O*s took part in each experiment. They were drawn more or less at random from a total group of 35 people.

The methods used were magnitude production, category production, and cross-modality matching. Examples of the form of the written instructions used may be found elsewhere (Stevens & Guirao, 1962). For lines the instructions required *O* to produce stimuli whose apparent longness (or shortness) was proportional to a series of numbers spoken in irregular order by *E*. A "standard" stimulus was presented first and called 10.

---

[1] This research was supported in part by Grant G-10716 from the National Science Foundation and in part by Grant B-2974 from the National Institutes of Health (Laboratory of Psychophysics Report PAR-275).

[2] Holder of a fellowship from the Consejo Nacional de Investigaciones Científicas y Técnicas, Argentine Republic.

With the method of category production, both the largest stimulus (called 7) and the smallest (called 1) were presented at the outset.

When the lines were to be matched to loudness (or softness) and to brightness (or dimness), the written instructions were modeled on the following format, with appropriate modifications:

> I am first going to present two stimuli, a line and a noise. Assume that the length of the line matches the loudness of the noise. I will then present the noise at various levels. Your task is to adjust the length of the line so that its apparent longness matches the loudness of each noise.

The need to present examples of the "matching" stimuli—a particular line and a given level of the noise—is dictated by the need to stay on scale, for the line could be adjusted only over a range of about 200 to 1, which is close to the practical limit. If a much wider range were possible it would then be feasible and desirable to let *O* choose his own matching length, just as it is preferable to let *O* choose his own modulus when the method of magnitude estimation is used (Stevens, 1956).

When the line was adjusted to match loudness (or softness), *O* wore a pair of earphones (PDR-8) through which a band of white noise (500 to 5,000 cps) was presented at one or another level.

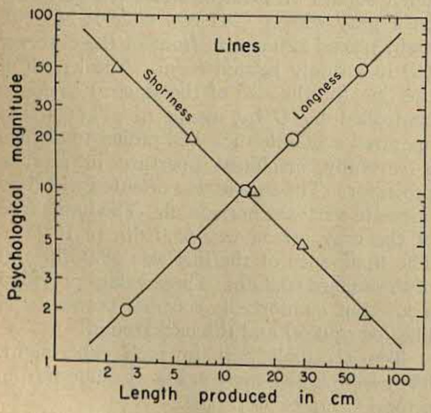


FIG. 1. Magnitude production with lines. (The points are the geometric means of the lengths produced in response to a series of numbers spoken by *E* to represent psychological magnitude. Each point is based on 20 judgments. The coordinates are logarithmic. The lines fitted by eye to the points have the slope, 1.0 for longness and .97 for shortness. The standard presented first and called 10 had a length of 13.5 cm.)

When the line was adjusted to match brightness (or dimness), *O* viewed a trans-illuminated target subtending a visual angle of about 3° and located on his left. Its luminance was varied by neutral filters. By turning his eyes *O* could also see and adjust the variable line, whose luminance in the darkened room was fixed at 58 db. re 10^{-10} L.

Geometric means were computed for the settings made in magnitude production and cross-modality matching. Arithmetic means were computed for the category productions. Each of the 10 *O*s made two adjustments for each level in an irregular order in all the experiments except those involving cross-modality matches, where three adjustments of the line were made for each level of the stimulus (light or sound).

RESULTS

Magnitude scales.—When *O*s judged the apparent magnitude of lines by adjusting their length to values proportional to numbers spoken in an irregular order, the lengths produced turned out to be almost exactly proportional to the numbers spoken. When the task was to make the apparent shortness proportional to the spoken numbers, the agreement was almost as close. The apparent shortness, therefore, grows as the reciprocal of the physical length. These relations are shown in Fig. 1, where, in log-log coordinates, the two slopes are close to plus 1.0 and minus 1.0. The two functions are therefore power functions with exponents equal to plus and minus 1.0, respectively.

Subjective shortness is thus seen to be the reciprocal of subjective longness. Even with this continuum, however, and with an arrangement that made the task extremely easy, there is a slight but detectable tendency for the inverse aspect, shortness, to give results that are concave downward in the log-log plot of Fig. 1. This behavior has been observed before (see Eisler, 1962; Stevens & Guirao, 1962) and is shown clearly in Fig. 8 and 9. There is evidence that a bias sometimes enters when a person is asked to judge in terms of the inverse of an attribute. In the present experiments the variability was generally greater for judgments of the inverse aspects, as is shown in Table 1.

The exponent for apparent length determined by these studies is slightly smaller than that found in previous experiments (Reese, Reese, Volkmann, & Corbin, 1953; Stevens & Galanter, 1957; Teghtsoonian, 1960). Even though all these various studies involved rather different procedures, they are all essentially agreed that length obeys the psychophysical power law and that the exponent is not far from unity. A line that is twice as long as another looks about twice as long.

TABLE 1

*SD*s of the Logarithms of the Lengths and of the Areas Produced in Response to Numbers Spoken by *E*

Length			Area		
No.	Longness	Shortness	No.	Largeness	Smallness
2	0.15	0.13	2.5	0.24	0.24
5	0.11	0.16	5	0.17	0.14
10	0.08	0.13	10	0.10	0.21
20	0.08	0.12	20	0.26	0.29
50	0.09	0.21	40	0.24	0.25
			80	0.28	0.31
Average	0.10	0.15	Average	0.21	0.24

Note.—Each entry is based on two judgments by each of 10 *O*s.

When the stimulus is a square area, however, the judgments of apparent size do not show a simple proportionality to the physical stimulus. Figure 2 shows the results for judgments of the largeness and smallness of squares. The data fit the power law and show a reciprocality between the two inverse aspects, but the exponent has the value of .7. This means that a

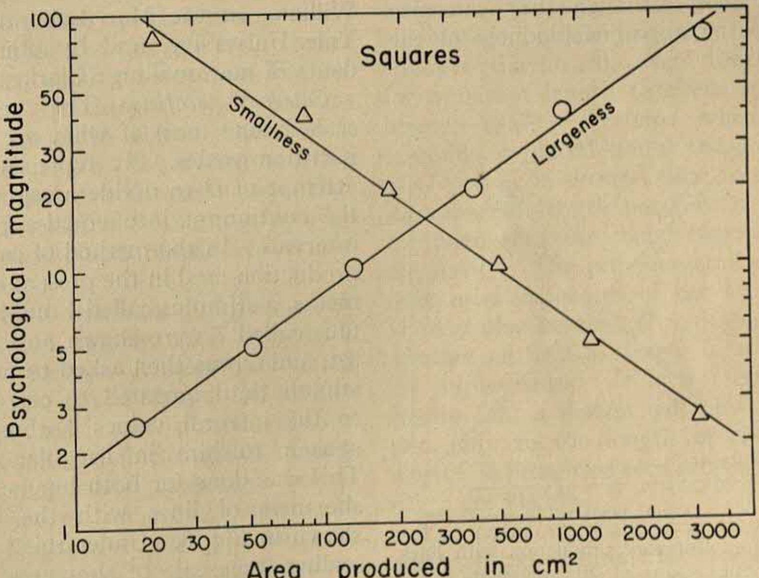


FIG. 2. Magnitude production with squares. (Geometric means of 20 judgments, in logarithmic coordinates. The slope for largeness is .67 and for smallness .72. The standard presented first and called 10 had an area of 100 cm.² for largeness and 441 cm.² for smallness.)

square that has twice the area of another does not look twice as large.

The standard deviations for the judgments of largeness and smallness are shown in Table 1.

It is possible, of course, to instruct *O* to make his judgments conform, not to the apparent size, but to the physical area. The value of the exponent may then rise to approximately 1.0, as was shown by Teghtsoonian (1960). She used figures cut from white cardboard and *O*s made numerical estimations of either the apparent size or the physical area of several kinds of figures. For example, with circular stimuli the exponent was .73 when *O*s judged apparent size, and 1.03 when they judged physical area. This is a demonstration of the fact that an *O* can often, as he pleases, either judge his sensations or commit the "stimulus error." In like manner, some acoustical engineers learn to judge sounds in terms of the standard scale of decibels, although they can also report the apparent loudness of the sounds.

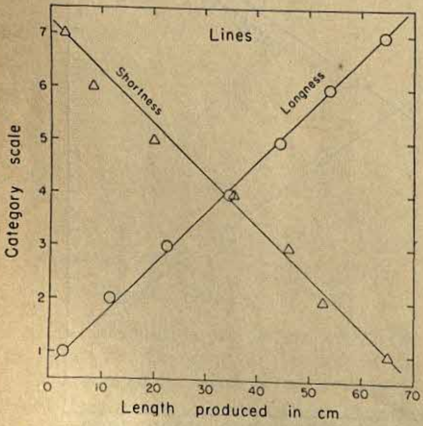


FIG. 3. Category production with lines. (Arithmetic means of 20 judgments. The values presented first and called 1 and 7 were 2.6 cm. and 64 cm. for longness and the reverse for shortness.)

When the exponent happens to have a value close to 1.0, as with lines, it is not easy to tell whether *O* is committing the stimulus error. But the existence of clear cases in which both attitudes can be taken by *O*, with correspondingly different functions produced, argues against the notion that all judgments of sensory magnitude are based on *O*'s knowledge of a stimulus correlate.

The exponent for the apparent size of areas has certain practical implications for the science of cartography, where symbols of different sizes are used to convey information of one kind or another. Given that the apparent size ψ is related to the area A by the formula $\psi = kA^{0.7}$, it follows that for one symbol to appear twice as large as another its area would need to be about 2.7 times as great. Its diameter, therefore, would need to be only about 1.6 times as great. This general principle has been discovered and utilized by R. L. Williams at the Map Laboratory of Yale University, and by other students of map making (Clarke, 1959).

Category scaling.—The category scale is one form of what are called partition scales. It represents the attempt of *O* to divide a segment of the continuum into equal-appearing intervals. In the method of category production used in the present experiments, a stimulus called 1 and a stimulus called 7 were shown at the outset, and *O* was then asked to produce stimuli that appeared to correspond to the integral values 1 through 7 spoken to him in irregular order. This was done for both longness and shortness of lines with the results shown in Fig. 3. Under this kind of scaling, the scale of shortness is the complement, not the reciprocal, of the scale of longness. In other words, the category scale for shortness is

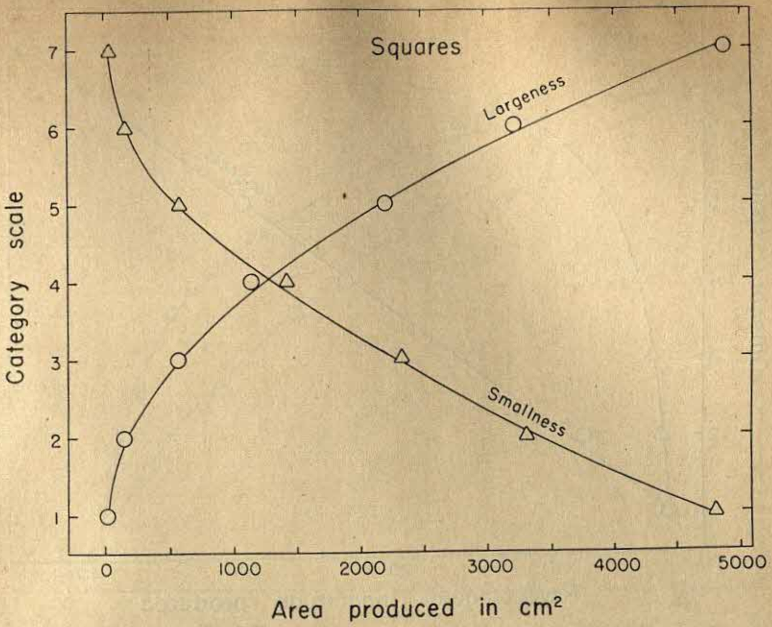


FIG. 4. Category production with squares. (Arithmetic means of 20 judgments. The values presented first and called 1 and 7 were 25 cm.^{2} and 5,041 cm.^{2} for largeness and the reverse for smallness.)

the same as the category scale for longness turned upside down.

Similar procedures applied to areas gave the results shown in Fig. 4. Whereas the category scale for apparent length is almost linear with stimulus length, the category scale for area is decidedly nonlinear. This is as it should be if the exponent for the apparent size of squares is less than 1.0.

In Fig. 3 and 4 the category scales are plotted against their respective physical scales. A more interesting comparison can be made if the various category scales are plotted against their respective linear scales of subjective magnitude, as is done below.

Category vs. magnitude scales.—In Fig. 5, the category scales for longness and shortness are plotted against their respective subjective scales (derived from Fig. 1). It turns out that the category scale for longness is almost linear against the subjective scale of longness. On a prothetic continuum the category scale is nearly always concave downward, even the scale for apparent length (Stevens & Galanter, 1957). Perhaps when the stimulus is placed under the control of *O* and he is allowed time to study it, apparent length becomes so easy to judge that the usual bias is not apparent. (The previous study used only brief exposures of the stimuli.) It may also be that *O* makes use of *position* as well as *length* in making his adjustments. It is not easy to ensure that a person will pay attention only to the length of the line when it is presented on a screen whose edges are visible.

The category scale of shortness is also plotted against the subjective scale of shortness in Fig. 5. The curve is a hyperbola. (The errant point corresponds to the value for Category

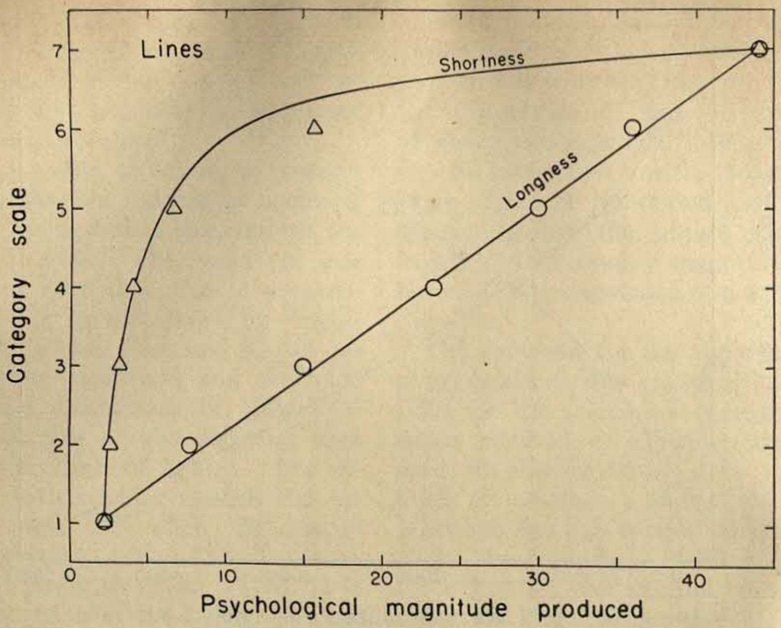


FIG. 5. Category scales plotted against their respective magnitude scales derived from Fig. 1. (Since the category scale for smallness is almost linear and the magnitude scale is a reciprocal function, the curve for smallness approximates a hyperbola. The abscissa represents longness and shortness, both in subjective units.)

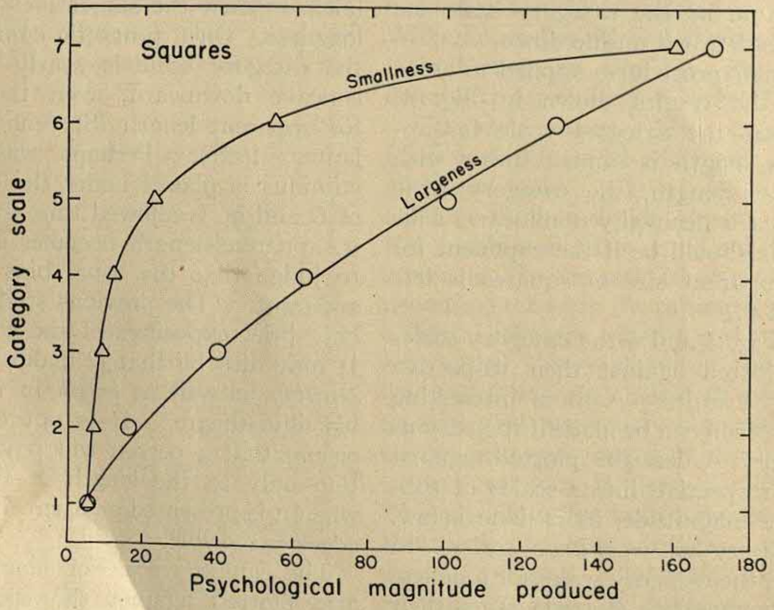


FIG. 6. Category scales plotted against their respective magnitude scales derived from Fig. 2. (The curvature that is typical of category scales on prothetic continua is evident in the curve for largeness. The abscissa represents largeness and smallness, both in subjective units.)

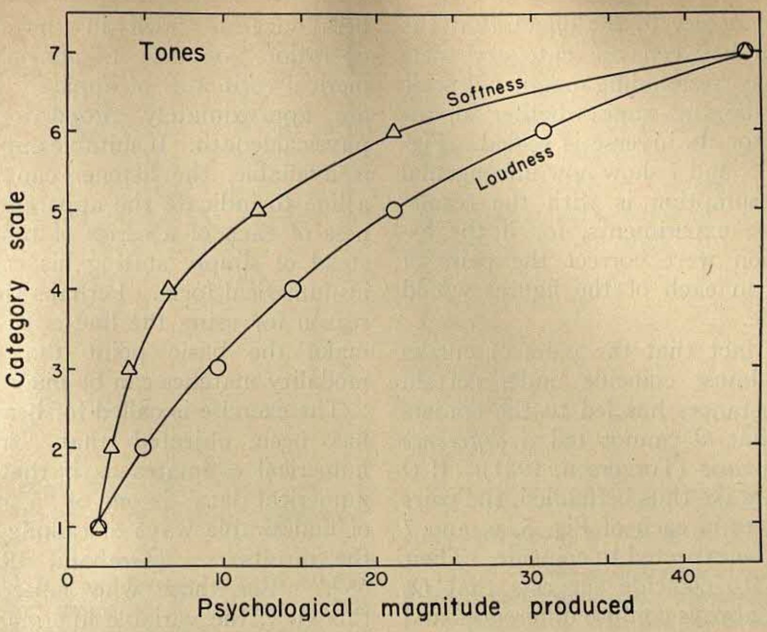


FIG. 7. Category scales for loudness and softness plotted against their respective magnitude scales. (The data are from an earlier set of experiments. The abscissa represents loudness and softness, both in subjective units.)

6 in Fig. 3.) In order to understand the necessity for this hyperbolic relation, it is sufficient to recall that for longness and shortness the two category scales are linearly related (as complements) and the two magnitude scales are reciprocally related. Then if category longness vs. magnitude longness gives a straight line, category shortness vs. magnitude shortness must give a hyperbola.

When the category scale is not a linear function of the magnitude scale, the category scale of the inverse aspect does not ordinarily give a hyperbola. Thus in Fig. 6 the category scale for largeness of squares is slightly concave downward against the magnitude scale. Correspondingly the category scale for smallness shows slightly less than hyperbolic curvature.

In order to extend this principle with the help of another example, comparable curves for the loudness and softness of tones are shown in Fig. 7. The data are from Stevens and Guirao (1962) and were obtained by the methods of category and magnitude production. The category scale for loudness is more curved and the category scale for softness is less curved than are the comparable scales for lines and squares in Fig. 5 and 6. If the progression in scale form exhibited in Fig. 5, 6, and 7 could be continued, the two category scales would ultimately coincide. At that point they would both be logarithmic functions of their respective magnitude scales, as Eisler (1962) says they should be. It appears, however, that when proper precautions are taken to avoid biases in the category scale the logarithmic form is never reached (see Stevens & Galanter, 1957). This fact bears on the provocative assumption enunciated by Eisler to the effect that the relation between the category scale and its corresponding magnitude scale should be the same whether an attribute or its inverse is scaled. Figures 5, 6, and 7 show how uncongenial this assumption is with the results of these experiments, for if the assumption were correct the pairs of curves in each of the figures would coincide.

The fact that the pairs of curves may almost coincide under certain circumstances has led to the conjecture that *O* cannot tell a *difference* from a *ratio* (Torgerson, 1961). If *O* were always thus befuddled, the pairs of curves in each of Fig. 5, 6, and 7 would be expected to coincide. Their failure to coincide suggests that *O*s do not always confuse differences and ratios.

It is an interesting question why there is a progressively greater curvature in the category scale as we go from longness to largeness to loudness. Perhaps an answer lies in the difference in *O*'s precision when he makes judgments on these three continua. In the magnitude productions the average variability in setting length of lines was .10 log unit. In setting areas it was .21 log unit. In setting tones it was .40 log unit (of sound pressure). These average standard deviations support a suggestion made previously (Stevens & Guirao, 1962) to the effect that the amount of curvature in partition scales is related to the amount of "noise" or variability present. This seems to be true on prothetic continua, where the subjective jnd increases with magnitude. It does not apply to metathetic continua where the subjective jnd tends to be constant over the continuum.

Cross-modality comparisons.—The use of a line as a substitute for numbers suggests itself in the scaling operation, because a person's numerical estimates of apparent length are approximately proportional to physical length. If suitable apparatus is available, the listener can adjust a line to indicate the apparent loudness of each of a series of tones, instead of simply stating his estimate in numerical form. Perhaps the best reason for using the line is to try to make the basic point that cross-modality matches can be made.

The exercise is called for because it has been objected that "treating numerical estimates as if they were numerical data" is one of "a number of undesirable ways of talking about the results . . . [Graham, 1958, p. 68]." For those who subscribe to this view, the variable line may offer a plausible alternative. Since any continuum can be matched by *O* to any other continuum, the line is not the only alternative, but it suggests itself as a "criterion continuum" because it is relatively easy to work with and its own exponent appears to be close to 1.0.

The results of an attempt to match lines to the loudness and softness of noise are shown in Fig. 8. The outcome is almost identical to that obtained when people judge loudness and softness by magnitude estimation. Several features are evident: the psychophysical power law is verified, the slope for softness is approximately equal in magnitude but opposite in sign to the slope for loudness, the points for softness show a slight downward concavity, and the slope (exponent) for loudness is .69, which is well within the range usually found for judgments of a noise of this kind.

A similar experiment in which *O*s matched length of lines to brightness and dimness gave the results shown in Fig. 9. Here again the power law

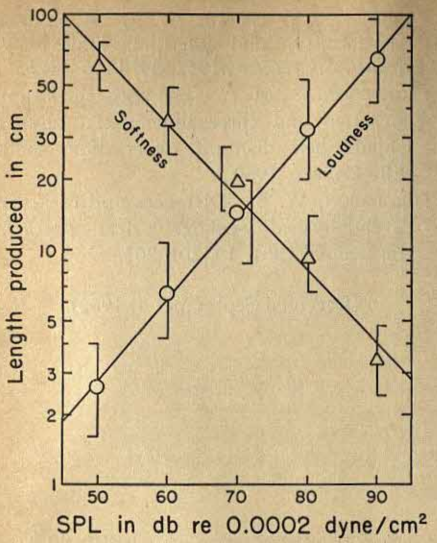


FIG. 8. Cross-modality matches between a variable line and levels of a band of noise. (Geometric means of 30 judgments. The first pair presented to define a "match" for loudness was 80 db. and 20 cm.; for softness 70 db. and 20 cm. Vertical bars show interquartile ranges.)

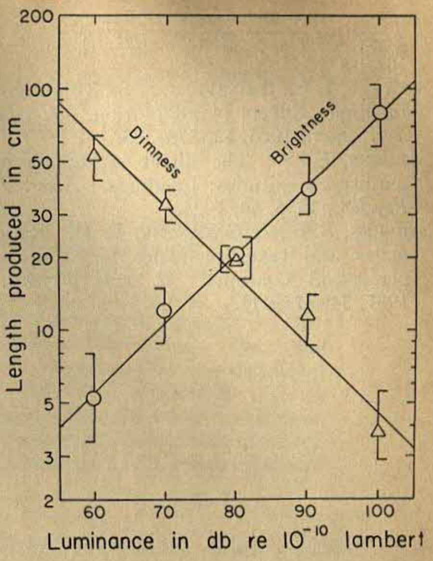


FIG. 9. Cross-modality matches between a length and a luminance. (Geometric means of 30 judgments. The first pair presented to define a "match" was 80 db. and 20 cm. for both brightness and dimness. Vertical bars show interquartile ranges.)

is verified and the approximate reciprocity between the attribute and its inverse is shown to exist. The exponent for brightness turns out to be .3, which is close to the consensus of many experiments with magnitude estimation and other methods (Stevens & Stevens, 1963).

The interquartile ranges shown by the vertical bars in Fig. 8 and 9 deserve comment. The variability is less when length is matched to brightness than when it is matched to loudness. The averages of the logarithmic interquartile ranges are for brightness .25 log unit, for loudness .39 log unit. On the other hand, the matching of length to dimness and softness gave average interquartile ranges of .18 and .29 log unit. In these experiments the cross-modality match to the inverse aspect gave the smaller variability. As noted above, the variability turned out to be greater when the inverse aspects of length and area were estimated numerically.

Taken all together, the results portrayed in Fig. 8 and 9 show that the psychophysical power law can, if need be, stand solely upon cross-modality matches, with no appeal made to subjective numerical estimates of any kind. Most of what we would lose by this gambit would be convenience.

REFERENCES

CLARKE, J. I. Statistical map reading. *J. Geogr. Ass.*, 1959, **44**, 96–104.

EISLER, H. Empirical test of a model relating magnitude and category scales. *Scand. J. Psychol.*, 1962, **3**, 88–96.

GRAHAM, C. H. Sensation and perception in an objective psychology. *Psychol. Rev.*, 1958, **65**, 65–76.

REESE, E. P., REESE, T. W., VOLKMANN, J., & CORBIN, H. H. (Eds.) Psychophysical research: Summary report 1946–52. *USN Spec. Devices. Cent. tech. Rep.*, 1953, No. 131-1-5.

STEVENS, J. C., & STEVENS, S. S. Brightness function: Effects of adaptation. *J. Opt. Soc. Amer.*, 1963, **53**, 375–385.

STEVENS, S. S. The direct estimation of sensory magnitudes: Loudness. *Amer. J. Psychol.*, 1956, 69, 1–25.

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

STEVENS, S. S., & GUIRAO, M. Loudness, reciprocality, and partition scales. *J. Acoust. Soc. Amer.*, 1962, **34**, 1466–1471.

TEGHTSOONIAN, M. A. The apparent size of one, two and three dimensional figures. Unpublished doctoral dissertation, Radcliffe College, 1960.

TORGERSON, W. S. Distances and ratios in psychophysical scaling. *Acta psychol., Amsterdam*, 1961, **19**, 201–205.

(Received September 4, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 187–194

EVENT UNCERTAINTY, PSYCHOLOGICAL REFRACTORY PERIOD, AND HUMAN DATA PROCESSING[1]

LYLE R. CREAMER[2]

University of Illinois

An experiment was performed to (a) test the psychological refractory period (PRP) hypothesis of the human S as a 1-channel data processing system with an event uncertain-time certain stimulus series, and (b) determine the combined effect of time and event uncertainty. To test the PRP hypothesis, 5 groups of 12 Ss had 5 different time intervals between visual and audio stimuli (0, 100, 200, 400, and 800 msec.). The visual S-R sequence produced a delay in audio responses that was significant at 0-, 100-, and 200-msec. intervals. The human apparently responds as a 1-channel system when he processes event uncertain-time certain data. Another group of 12 Ss had the 5 interstimulus intervals with both time and event uncertainty. Their bisensory audio impairment was similar to the 5 time certain groups.

The question of the limits of human data processing has stimulated a number of studies under the rubric of psychological refractory period (PRP) (e.g., Davis, 1956, 1957, 1959; Fraisse, 1957; Hick, 1948; Telford, 1931; Vince, 1949; Welford, 1959). On the basis of their data, these investigators argue that the human responds as a one-channel system because his central processes are refractory during an S-R sequence. If a second stimulus arrives during the refractory period, the response to it will be delayed until the central channel has recovered from the first S-R sequence. Delay is inferred from response to a control stimulus series that does not have a first S-R sequence to occupy the single channel.

An alternative interpretation of this delay for the second response has been offered by Adams (1962), Elithorn and Lawrence (1955), and Poulton (1950). In an experimental situation where events are certain as to their type, they contend that the delay for the second response is not due to PRP; but to a lack of preparedness for the second stimulus because of inappropriate temporal expectancies. The *S* has learned to expect a longer temporal interval between stimulus pairs on the average, and impairment is evidenced because he is not set to respond when a short interval occurs. The contrast between the PRP and the expectancy view is clear. The PRP is an intrinsic central delay independent of the temporal patterning of events, and expectancy holds the decrement in the second response to be a by-product of temporal learning in a situation where event type is certain but time intervals are uncertain. Given enough event and temporal certainty, *S* can be a multichannel system processing two or more stimulus streams concurrently and without impairment, according to the expectancy position.

With the exception of Adams and Chambers (1962) and Adams (1962),

---

[1] Based on a dissertation submitted in partial fulfillment of the requirements for a PhD degree in psychology at the University of Illinois. The thesis was under the direction of Jack A. Adams.

This research was supported by the United States Air Force under Contract AF 49(638)-371, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

[2] Now at Long Beach State College.

experiments on this topic have not discriminated between these two points of view. With event certainty in concurrent audio and visual stimulus series, Adams (1962) found that the decrement in response to the second of two closely spaced stimuli was a function of the statistical structure of the time intervals between the bisensory stimulus pairs. On the other hand, Adams and Chambers (1962) found decrement under conditions of complete temporal certainty, providing there was uncertainty about the type of event that was to occur next. It would seem from the Adams-Chambers study that the basic defining conditions for PRP are linked to event uncertainty. The experiment reported here exploits this lead by Adams and Chambers and seeks to establish the nature of the delay function for response to a second stimulus when events in concurrent bisensory stimulus series have event uncertainty.

METHOD

Purpose of the Experiment

The purpose of the main experiment was to test the hypothesis that event uncertainty in two concurrent stimulus series is a basic defining condition for the limit of human data processing. Complete temporal certainty is used throughout so that all *S*s can fully learn when events will happen—thus eliminating differential temporal expectancies as a determiner of impairment. One stimulus was visual for response by one hand and the other stimulus was audio for response by the other hand. The occurrence of each stimulus was uncertain as to type (one of two possibilities). Five groups of *S*s were used, where a group always had either 0, 100, 200, 400, or 800 msec. between the two uncertain visual and audio stimulus events each time they were presented. Each group provided its own unisensory control measure for determining impairment in bisensory responses. It is hypothesized that decrement in response to the second stimulus will be maximal at the zero interval, and decrease as the interstimulus interval increases.

A second aspect of the research is embodied in a single group and asks how impairment is a compound function of event and temporal uncertainty. This group had the same event uncertainty conditions as the other groups and, in addition, had time uncertainty. An *S* could neither predict what stimulus would occur in each sensory dimension nor when it would occur. This group also provided its own unisensory control measures.

Apparatus

The Discrete Tracking Apparatus previously described by Adams and Chambers (1962) and Adams (1962) was used with slight modifications to present three visual and three audio stimuli. Basically, the apparatus included a Friden motorized digital tape reader, three No. 47 General Electric lamps, three Hewlett-Packard audio oscillators, and an Esterline-Angus operations recorder. The lights, tones, and rest periods were programed on punched tape which was read by the motorized tape reader at a rate of 10 holes per second. After a visual stimulus was read, 14–30 msec. elapsed before the lamp glow could be seen by *S*. The lag for the audio stimuli was 7–10 msec. The three stimulus lights were mounted in a 2 × 3¼ in. box which was 30 in. in front of the seated *S* at eye level. Each light was mounted behind a jeweled filter (red, white, or green) and a neon feedback lamp was placed directly below each filter. Responses were indicated by the feedback lamps and a lighted feedback lamp beneath a lighted stimulus lamp signaled a correct response. An incorrect response was indicated when no feedback lamp was on or when the lighted feedback lamp was not under the lighted stimulus lamp. The three tones (low, 400 cps; medium, 700 cps; and high, 1,000 cps) were presented through earphones and pure or complex tones provided feedback. A correct response produced a pure tone while an incorrect response produced a complex tone. The complex tone was a combination of two of the three frequencies; the correct stimulus tone was one of the two tones and superimposed on it was a rapidly interrupted tone for the position of the incorrect response. When no button was depressed *S* received a pure stimulus tone so *S* had to make a response to get feedback. The visual and audio displays directly presented both stimulus and response information, making them pursuit tracking tasks. Three button-type switches 1½ in. in diameter were arranged in an equilateral triangle on each arm of *S*'s chair and there were 2½ in. between the centers of the buttons. The *S* responded by depressing the buttons on one armrest for the visual stimuli and the buttons on the other armrest for the audio stimuli. Each button on an armrest corresponded to one of the three stimuli in a sensory dimension. The size and the arrangement of the buttons allowed *S* to respond without visually checking his hand movement.

Procedure

Six groups of *S*s were used—one group of 12 *S*s for each of the five interstimulus intervals (0, 100, 200, 400, and 800 msec.) with time certainty (TC) and one group of 12 *S*s served in a time uncertain (TU) interstimulus interval condition. All *S*s received Trials 1–21 on Day 1 and Trials 22–46 on Day 2, and each trial was 93 sec. long. No less than 24 hr. and no more than 1 wk. separated the 2 experimental days. On each day, all trials were separated by 30-sec. rest periods with the exception of a 5-min. rest following Trials 12 and 24. A warning light preceded each trial for 5 sec.

Groups TC.—Unisensory visual, unisensory audio, and bisensory visual-audio tracking tasks were used. In unisensory visual tracking, *S* had to press the response buttons and keep the feedback lamp vertically aligned with the lighted stimulus lamp. In unisensory audio tracking *S* had to press the correct buttons and maintain a pure tone. Bisensory trials had both visual and audio tracking at the same time where the tone change occurred with or slightly after the light change. The interval between the light change and the tone change had previously been described as the interstimulus interval.

On Day 1 each *S* received information about interstimulus intervals and appropriate responses. The *S*s were told that if the overlapping visual and audio resolution processes caused interference so that they could not respond to both changes as they occurred, they should respond to the visual change first and then follow this, as quickly as possible, with their audio response. The *S*s were encouraged to avoid a delayed response for both stimuli if this was possible. By random assignment, half of the *S*s used the right-hand buttons for the visual responses and left-hand buttons for the audio.

Trial 1 was a unisensory visual condition where *S*s had to respond only to the light changes and no tones were presented. Trial 2 was a unisensory audio condition where only tones were presented. Trials 3, 4, 5, and 6 were bisensory visual-audio. The remaining trials repeated this cycle except for the eighth repetition where the last two bisensory trials had to be omitted because of time limitations.

In the unisensory conditions, a trial had 30 discrete events of 3 sec. duration each. Each trial started with a standard initial stimulus (white for the lights and medium for the tones) and then the stimuli were selected so that after any stimulus (e.g., green) the other two stimuli within that modality (e.g., red and white) occurred equally often in a random order. For example, a high tone was followed equally often by a low and a medium tone and the order was randomly determined. No stimulus ever directly followed itself. In a bisensory condition, after the initial stimulus, a trial had 30 visual and 30 audio events made up by overlapping the two unisensory conditions; thus, there was 3 sec. between all light changes and 3 sec. between all tone changes but the tone changes followed the light changes by one of the five interstimulus intervals defining the five TC groups.

The Operations Recorder was used on Trials 43, 44, and 45 to obtain a detailed graphic record of stimuli and responses.

Group TU.—The procedure for this group was the same as that used for TC groups with three exceptions: (*a*) The stimulus series was changed so that there was always time uncertainty. The unisensory trials had the standard initial stimulus followed by 30 discrete events. The duration of the events varied from 2.6 to 3.4 sec. at .1-sec. intervals; but, they had an average of 3 sec. duration. Each bisensory trial had all of the five interstimulus intervals that were used in the TC task. Each interval occurred six times per trial and the order was randomly determined. (*b*) The instructions were altered to describe the temporal uncertainty associated with the stimuli. (*c*) The Operations Recorder was used on Bisensory Trials 39, 40, and 46, in addition to Trials 43, 44, and 45.

Subjects

Seventy-two male undergraduates were randomly assigned to six experimental treatments (12 *S*s per group) with one exception. Group TU was not started until 80% of the other conditions had been completed. The TU *S*s were spaced randomly among the remaining TC *S*s. The *S*s were paid for their participation.

Performance Measurement

Response speed (RS) and frequency of errors were used as response measures for both visual and audio data. Response speed was the primary response measure and it was used as a measure of the time interval after the stimulus until the correct response was evoked. For RSs, the tracking task was treated as a series of two-choice reaction time problems where each problem served as the warning for the next problem and where each response served as the starting position for the next problem. Bisensory tracking was a dual two-choice reaction time problem where the audio problem occurred with or slightly after the visual problem. An RS score for *S* in the TC groups was the median of 30 responses in a response class on a trial. Unisensory RSs for the TU group were computed just like the RSs for the TC groups; however, bisensory visual and bisensory audio RSs for Group TU were the median of 24 responses in a response class on four trials. Performance on the additional trials was tested statistically and not found to differ significantly from performance on Bisensory Trial 45. Response speeds were recorded on Trials 43 (unisensory visual), 44 (unisensory audio), and 45 (bisensory visual-audio) for the TC and TU groups and RSs were also recorded on Bisensory Trials 39, 40, and 46 for Group TU. The RSs were read to the nearest one-half chart unit, where one-half chart unit was equal to 33 msec. Statistical tests used chart units but the RSs were converted to milliseconds for tables and graphs.

The frequency of incorrect responses was the second performance measure. The frequency of unisensory visual (Trial 43), unisensory audio (Trial 44), bisensory visual (Trial 45), and bisensory audio (Trial 45) errors was obtained for the TC groups and statistical tests. The percentage of unisensory visual (Trial 43), unisensory audio (Trial 44), bisensory visual (Trials 39, 40, 45, and 46), and bisensory audio (Trials 39, 40, 45, and 46) errors was obtained for Group TU and statistical tests and graphs used percentage of erroneous responses because unisensory and bisensory scores were based on a different number of responses. Each *S* in Group TU had a unisensory visual, a unisensory audio, five bisensory visual, and five bisensory audio error scores (one bisensory score for each of the five interstimulus intervals).

TABLE 1

MEAN AUDIO RS SCORES IN MSEC.

Cond.	Interstimulus Interval (msec.)				
	0	100	200	400	800
TC Unisensory	671	585	647	703	740
TC Bisensory	902	773	743	697	697
TU Unisensory	663	663	663	663	663
TU Bisensory	958	911	824	766	692

Less than 1% of the visual responses for TC and TU groups were erroneous, so no evaluation of them was made.

RESULTS

TC Groups, Audio Performance

Response speed (*RS*).—The PRP hypothesis was tested by comparing an *S*'s bisensory audio RS with his unisensory audio RS. If the human organism is a one-channel data processing system that can accommodate only one uncertain event at a time, the prior S-R visual sequence would occupy the one channel and induce an impairment in a closely following S-R audio sequence. Larger bisensory audio delays are expected for the shorter interstimulus intervals, and delays are expected to diminish as the interval is increased.

Each *S* served as his own control so it was necessary to insure that the unisensory control measures did not differ among the groups. The unisensory audio RS score (Trial 44) for each *S* was entered in a simple analysis of variance; $F\ (4, 55) = .630$ was not significant at the .05 level. The .05 probability level was used for all significance tests.

An RS decrement score was obtained for each *S* by subtracting his unisensory audio RS (Trial 44) from his bisensory audio RS score (Trial 45). The audio RSs for Trials 44 and 45 are shown in Table 1 and the decrement scores, which directly reflect the degree of PRP impairment,

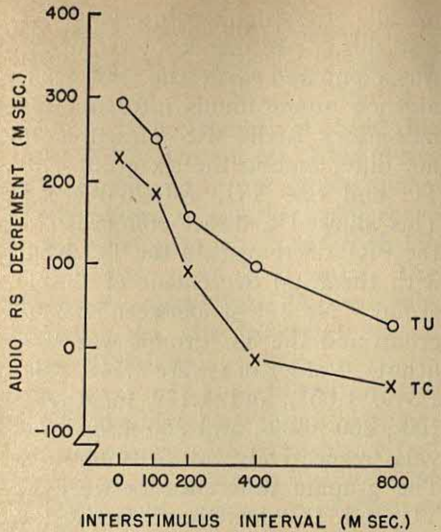


FIG. 1. Bisensory audio RS minus unisensory audio RS for TC and TU conditions.

are plotted in Fig. 1. The RS decrement for Group TU is also plotted in Fig. 1 and will be discussed subsequently. A simple analysis of variance was used to determine whether the decrement varied as a function of the interstimulus interval. The F (4, 55) = 8.566 was significant, and it can be seen in Fig. 1 that the greatest impairment occurred at the shortest interval and then impairment diminished to near zero at the 400-msec. interval. A t test for related measures was used for each interval to determine whether the decrement was significantly greater than zero. Only the t's for the 0-, 100-, and 200-msec. intervals, 8.075, 6.905, and 2.680, respectively ($df = 11$ for each), were significant. The t values for the 400- and 800-msec. intervals were .096 and .937, respectively, each with $df = 11$, and bisensory performance was slightly superior to unisensory performance for these two intervals giving a negative decrement score.

Errors.—Error scores were recorded to determine whether bisensory audio impairment was partially due to an increase of erroneous responses from unisensory to bisensory tracking, and whether the error impairment, if any appeared, was a function of the interstimulus interval.

The statistical treatment of error scores was similar to that used for RS scores. Unisensory audio errors (Trial 44) were tested with a simple analysis of variance among the TC groups for a difference in unisensory responding and F (4, 55) = 2.1466 was not significant. A decrement score was obtained by subtracting the unisensory audio errors from the bisensory audio errors (Trial 45 minus Trial 44) for each S and these error decrement scores are shown in Fig. 2 along with the TU data. A simple analysis of variance indicated no difference among the intervals in the degree of impairment, F (4, 55) = .849. The t tests for related measures at the five intervals were all nonsignificant indicating that no bisensory audio impairment resulted from the preceding visual S-R sequence.

Only 10% of all audio responses

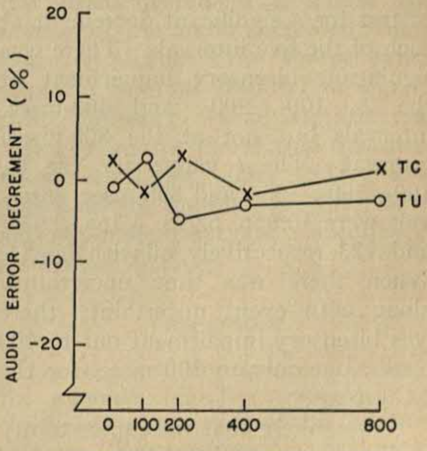


FIG. 2. Bisensory audio error scores minus unisensory audio error scores for TC and TU conditions.

were incorrect and the frequency of errors was fairly evenly divided between the unisensory and bisensory conditions.

Group TU, Audio Performance

Response speed.—To determine whether there was bisensory audio impairment, TU decrement scores for the five intervals were obtained by subtracting the unisensory audio RS (Trial 44) from the bisensory audio RS, and the decrement scores are plotted in Fig. 1. The unisensory and bisensory audio RSs are given in Table 1. The TU decrement scores for each of the five intervals were cast into an analysis of variance where the main effects were *S*s and Intervals, and these main effects were tested with the *S*s × Intervals interaction term. This test was used to determine whether the bisensory impairment varied as a function of the interval. The analysis for intervals, F (4, 44) = 33.395, was significant and the function resembled the TC function except that the impairment was slightly greater overall. A t test for related measures was used to test for a significant decrement at each of the five intervals. There was significant bisensory impairment at the 0-, 100-, 200-, and 400-msec. intervals but not at the 800-msec. interval. The t values for the 0-, 100-, 200-, 400-, and 800-msec. intervals were 10.066, 6.995, 4.766, 2.951, and .723, respectively, all with $df = 11$. When there was time uncertainty along with event uncertainty there was bisensory impairment out to 400 msec., but only to 200 msec. for the TC groups.

It was likely that time uncertainty would cause additional bisensory audio impairment so that TU decrement scores would be significantly greater than TC decrement scores for some or all interstimulus intervals. To test this, performance of Group TU was compared with the TC groups at each interstimulus interval. The unisensory audio RS measures did not differ among the six groups, five TC and one TU, F (5, 66) = .534. This allowed a direct comparison of the PRP decrement of the TC groups with the total decrement of the TU group. No t test between the TU group and the TC groups was significant. Values of t were 1.548, 1.332, 1.349, 1.607, and 1.179 for the 0-, 100-, 200-, 400-, and 800-msec. intervals, respectively, all with $df = 22$. The graphic difference between the TU and TC decrement received no statistical support.

Errors.—The statistical treatment of errors was similar to that used for RS scores. Decrement scores were obtained by subtracting the percentage of unisensory errors from the percentage of bisensory audio errors for each interstimulus interval. These decrement scores are plotted in Fig. 2 along with the TC function. Using the decrement scores, an analysis of variance was used to determine whether the error decrement was a function of the interstimulus interval. The test had *S*s and Intervals as main effects and these were tested with the *S*s × Intervals interaction. The F (4, 44) = 3.253 was significant. A t test for related measures was used to determine which interval resulted in bisensory impairment. The t values for the 0-, 100-, 200-, 400-, and 800-msec. intervals were .323, .841, 2.464, 1.894, and 1.419, respectively (each with $df = 11$), and only the t ratio at the 200-msec. interval was significant. At 200 msec. the decrement score was negative so bisensory performance was actually better than unisensory performance in terms of errors.

TC Groups, Visual Performance

Response speed.—The statistical treatment of the visual RSs was similar to the treatment used for audio RSs. Unisensory RSs (Trial 43) were entered into a simple analysis of variance to test for a difference among the groups. The F (4, 55) = .722 was not significant. The unisensory scores (Trial 43) were subtracted from the bisensory visual scores (Trial 45), yielding a decrement score which reflected impairment of the bisensory visual response. The unisensory and bisensory visual RS means are given in Table 2 and the decrement scores are plotted in Fig. 3 along with the TU decrement. The simple analysis of variance for the decrement scores yielded a significant F ratio, F (4, 55) = 3.329. The t tests for related measures to determine which of the intervals resulted in impairment were all significant except for the 800-msec. interval. The t values for the 0-, 100-, 200-, 400-, and 800-msec. intervals were 2.490, 4.080, 4.171, 4.923, and 1.382, respectively, each with df = 11.

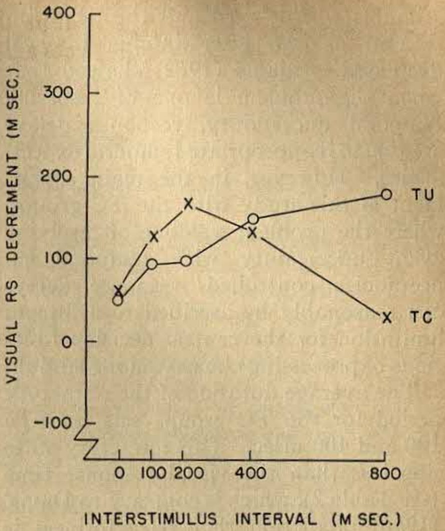


FIG. 3. Bisensory visual RS minus unisensory visual RS for TC and TU conditions.

Group TU, Visual Performance

Response speed.—Each S had one unisensory visual RS and five bisensory visual RSs, one for each interval. The decrement scores between the unisensory RSs and the bisensory visual RSs are plotted in Fig. 3 and the mean visual RSs are given in Table 2. The analysis of variance of the decrement scores had *S*s and Intervals as main effects and these were tested with the *S*s × Intervals interaction. The F (4, 44) = 7.494 was significant and it can be seen that the visual decrement is almost the reverse of the audio decrement function for Group TU. Every interstimulus interval resulted in significant bisensory visual impairment. The t values for the related measures tests were 2.290, 3.031, 3.532, 3.556, and 3.329 for the 0-, 100-, 200-, 400-, and 800-msec. intervals, respectively, all with df = 11.

TABLE 2

MEAN VISUAL RS SCORES IN MSEC.

Cond.	Interstimulus Interval (msec.)				
	0	100	200	400	800
TC Unisensory	502	467	447	460	508
TC Bisensory	566	606	618	605	544
TU Unisensory	453	453	453	453	453
TU Bisensory	510	553	558	606	636

DISCUSSION

The results of this experiment indicate that a human S can be considered a one-channel data processing system in accord with the PRP view providing that event uncertainty is present in the stimulus series. When an audio S-R sequence occurred with or closely followed a visual S-R sequence, response to the audio stimulus was delayed and the delay increased as the interstimulus interval decreased. Adams (1962) showed that when the problem is one of resolving temporal uncertainty, response delays are due to inappropriate temporal expectancies. However, in the main experiment of this study with the TC groups, where the problem was one of resolving event uncertainty with temporal expectancies controlled, response delays can reasonably be ascribed to a human limitation for the central decision functions of processing choices among stimuli.

The average duration of the refractory period for the TC groups was between 200 and 400 msec. This refractory state was less than the visual response time (see Table 2), which is contrary to Davis' (1956) prediction that refractoriness is as long as the reaction to the first stimulus. This minor tolerance for accommodating two uncertain S-R sequences could be viewed as an attenuation of a strict one-channel position, but the extensive impairment observed for the second response when the interstimulus intervals were brief argues convincingly for human intolerance of two closely spaced uncertain events.

Group TU had time uncertainty added to event uncertainty in the stimulus series, and significant bisensory audio impairment occurred from zero through 400 msec. delay. Their performance was comparable to the TC groups except that no impairment was found for the TC condition with 400 msec. delay. This was the only evidence that temporal uncertainty compounded the delay effects attributable to event uncertainty.

REFERENCES

ADAMS, J. A. Test of the hypothesis of psychological refractory period. *J. exp. Psychol.*, 1962, **64**, 280–287.

ADAMS, J. A., & CHAMBERS, R. W. Response to simultaneous stimulation of two sense modalities. *J. exp. Psychol.*, 1962, **63**, 198–206.

DAVIS, R. The limits of the "psychological refractory period." *Quart. J. exp. Psychol.*, 1956, **8**, 24–39.

DAVIS, R. The human operator as a single channel information system. *Quart. J. exp. Psychol.*, 1957, **9**, 119–129.

DAVIS, R. The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, 1959, **11**, 211–220.

ELITHORN, A., & LAWRENCE, C. Central inhibition: Some refractory observations. *Quart. J. exp. Psychol.*, 1955, **7**, 116–127.

FRAISSE, P. La période réfractaire psychologique. *Année psychol.*, 1957, **57**, 315–328.

HICK, W. E. Reaction time for the amendment of a response. *Quart. J. exp. Psychol.*, 1948, **1**, 175–178.

POULTON, E. C. Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, 1950, **2**, 99–113.

TELFORD, C. W. The refractory phase of voluntary responses. *J. exp. Psychol.*, 1931, **14**, 1–35.

VINCE, M. A. Rapid response sequences and the psychological refractory period. *Brit. J. Psychol.*, 1949, **40**, 23–40.

WELFORD, A. T. Evidence of a single-channel decision mechanism limiting performance in a serial reaction task. *Quart. J. exp. Psychol.*, 1959, **11**, 193–210.

(Received September 7, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 195–200

A FIGURAL AFTEREFFECT PRODUCED BY A PHENOMENAL DICHOTOMY IN A UNIFORM CONTOUR[1]

BERNARD WEITZMAN

Creedmoor Institute for Psychobiologic Studies, Queens Village, New York

If the central processes responsible for figural aftereffects derive exclusively from peripheral events, then the phenomenal status of a contour should not affect the figural aftereffect it produces. If the central processes also modify the consequences of the peripheral events, an alteration of the figural aftereffect might result. 44 college student Ss were tested in a figural aftereffect displacement paradigm in which the inspection figure contained a centrally fixated uniform contour, phenomenally dichotomized into figure and ground halves by a biased center contour. Control conditions for prefixation judgments and for methodological artifacts were included. At better than the .05 level a differential aftereffect was shown to exist. The results were interpreted as indicating the necessity of examining the role played by central processes in contour perception.

The Köhler-Wallach (1944) monograph on figural aftereffects has given rise to a considerable body of experimentation and theoretical discussion. The explanation of figural aftereffects presented in this monograph (the satiation theory) is based upon a model of spread of excitation which is not confined to the neurons, but which permeates the neural tissue of the CNS. Though initiated by peripheral events, the spread of the pattern of excitation through the cortical region is independent of retinal processes. The phenomenal location of a stimulus contour may be predicted as a function of the cortical locus in which the pattern of excitation has its greatest density (Köhler & Wallach, 1944, p. 328). Figural aftereffects derive from the way in which cortical patterning of new peripheral stimulation is affected by the cortical remains of a previously created pattern of excitation.

An alternative explanation was offered by Osgood and Heyer (1952). Maintaining the assumption of the primacy of the neural pathways as determiners of perceptual events, i.e., without recourse to a novel cortical theory, they were able to construct a descriptive model which explained many of the findings reported and explained by Köhler and Wallach. In order to make their model work, Osgood and Heyer invoked the retinal "on-off" mechanisms described by Marshall and Talbot (1940) in their work on contour perception. Using the frequencies of on-off firings produced at various retinal loci by nystagmus eye movements when fixating a contour, Osgood and Heyer were able to derive the patterns of excitation required to produce figural aftereffects by assuming cortical registration of the patterns of retinal excitation.

---

[1] This paper is based upon a portion of a dissertation submitted to the Psychology Department of the Graduate Faculty of Political and Social Sciences of the New School for Social Research, in partial fulfillment of the requirements for the PhD degree. The author expresses thanks to Robin Lobeck for the suggestion which led to the development of the method employed in this experiment, to his committee member, Gustav Levine, for his inestimably helpful concern and criticism at every stage of the research, and to J. R. Whittier, Director of Psychiatric Research at the Creedmoor Institute for Psychobiologic Studies, for the patience and support which made possible the successful completion of the experimentation.

The present paper addresses itself to a question generated by the controversy outlined above. If Köhler and Wallach are correct in their *approach* to the problem (without reference to the detail of their particular model), then there are processes relevant to contour perception, and hence to the phenomena of figural aftereffects, which are specifically cortical in nature. While these processes are stimulated into activity by peripheral events, the resulting perceptions may only be adequately comprehended by directing concern to the nature and role of cortical functions. If Osgood and Heyer are correct in their *approach* (again without reference to the detail of their model), then the relevant cortical processes may be conceived as directly derivative from peripheral events. The question may be put as follows: May we, by analyzing the peripheral consequences of a particular stimulus configuration, e.g., a particular pattern of on-off firings, directly predict the resulting cortical events, or must we consider the cortex as modifying stimulus input according to laws of its own function?

The experiment to be reported attempts to investigate the above question by the creation of an experimental paradigm in which an objectively uniform contour is presented in such a way as to be phenomenally nonuniform. If it is possible to predict the effects of this contour by referring only to its peripheral distribution, its objective uniformity should preclude the possibility of demonstrating a differential effect for part of its extent. If, on the other hand, as hypothesized, the phenomenal divergence from objective uniformity results in the production of differential displacement in a figural aftereffect paradigm, the participation of cortical processes in contour perception would be indicated.

METHOD

Experimental materials.—Four inspection figures were constructed. They were drawn with black India ink on screens of white construction board, 15 in. on a side and pasted on beaver board. The screens, when presented to *S*, were placed in a frame of ½-in. aluminum angle stock, painted black. This frame provided a ½-in. black border on the vertical sides of the screen. The top of the screen was not enclosed and the bottom rested on a beaver board box which was the support for the aluminum frame. This box was also painted flat black. Inspection Fig. 1 consisted of a black cross with each of its two arms 1.5 mm. thick and 7 mm. long, set at 45° to the vertical. The center of the cross corresponded to the center of the screen. This figure was used in all experimental conditions for determination of the presatiation points of subjective equality for all *S*s. Inspection Fig. 2 consisted of a black square, 2 in. on a side, vertically bisected by an outline human profile, the upper lip of which coincided with the center of the square and with the center of the screen. This figure represents the end point of an extensive search, in pilot investigations, for a biased contour, ambiguous figure, in which *S*s could maintain one organization for the bulk of a 3-min. inspection period. The present figure, because it is not truly ambiguous, was found to induce relatively few spontaneous reversals. Thirty pilot *S*s reported "no reversals," or "almost no reversals." The figure, while being reversible, tends to facilitate unidirectional organization. In this figure the profile faced to the left, i.e., the face occupied the right side of the figure. All contours of the figure, including the center profile, were 2 mm. thick. The area under each side of the profile was 1,250 mm.^{2} as measured by construction of millimeter columns. Inspection Fig. 3 was a mirror image of Inspection Fig. 2. Inspection Fig. 4 was identical in all dimensions to Inspection Fig. 2 and 3 except that the square was vertically bisected by a straight line instead of by a profile. At the center of this line, in a position corresponding to the center of the square, was drawn a cross identical in all respects to that described for Inspection Fig. 1.

In Inspection Fig. 2 and 3 the bottom contour is objectively uniform and describes one side of a square. This square, however, is bisected by a biased contour. The result is that the objectively uniform contour is phenomenally divided into two halves: one half participates in the figural organization directed by the profile and the other half is involved in this organization as part of a "ground." Inspection Fig. 2 and 3 thus meet the conditions for the testing of the hypothesis, i.e., they present a uniform contour in such a fashion that one half of the contour participates in a figural organization and the other half does not. Inspection Fig. 4 was introduced to act as a control over a possible artifact in the findings, i.e., it might be that a moving test line (see description of test figure below) within a region which has been, to some degree, satiated would, in this paradigm, produce an artifact via the "distance paradox" (Köhler & Wallach, 1944, pp. 281–289).

A test figure was constructed consisting of a fixation point (a black cross) with two black dashes placed symmetrically below it. The relationship of the dashes to the cross was such that the dashes fell immediately below the retinal location of the bottom contour of Inspection Fig. 2, 3, and 4, when the cross was at the retinal location of the fixation points of the inspection figures. The dashes were 1.5 mm. thick and each was 13 mm. long. They were placed 1.25 in. below the center of the cross (measured from their upper edges) and were centered about a hypothetical vertical projection from the center of the cross. The horizontal distance between the two dashes was 12 mm. The test figure was contained in an apparatus which permitted one of the dashes to be moved up or down. The cross and the dash which was to be nonmoving were engraved in a ⅜-in. thick sheet of Plexiglas and colored with black India ink. This sheet of Plexiglas was securely seated in an aluminum frame mounted on an aluminum box. Both frame and box were finished with a flat black paint. The movable dash was drawn in black India ink on a screen of white construction board pasted on beaver board. This screen was mounted in an aluminum frame which held it pressed against the engraved side of the Plexiglas.

The frame holding the white screen was mounted, top and bottom, through four aluminum posts, and ran in a track formed by the sides of the frame holding the Plexiglas. A fifth post, placed in the vertical axis of the movable frame, was machine-threaded through a 1-in. aluminum block which was fixed to the aluminum frame and through which the other four posts were free running. The white screen was moved up and down by turning a ratchet handle fixed to the top of the fifth post. A scale, marked in. 5-mm. intervals and marked with a "zero" position to indicate objective alignment of the two dashes, was attached to the stationary frame. A black border was provided for the test field with black masking tape, 2⅜ in. wide fixed to the surface of the Plexiglas. The field thus outlined was 15 in. on a side.

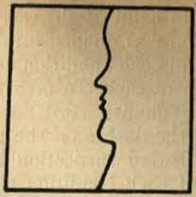


FIG. 1. Example inspection figure; Inspection Fig. 2.

As measured by a Photovolt electronic photometer, Model 501-M, at 6 ft. from the screens, the illumination of the inspection screen was .87 ft-c and of the test screen was .75 ft-c. This apparatus enabled *E* to displace one of the dashes closer to, or further from the retinal location of the inspection contour. The test figure was also equipped with a movable black card by raising of which *E* could cover the dashes while leaving the fixation point exposed, and by lowering of which the entire test figure could be exposed.

Subjects.—Forty-four college student volunteers were the *S*s in this experiment: 15 in Cond. A, 14 in Cond. B, and 15 in Cond. C. There were 16 male *S*s and 28 female *S*s, ranging in age from 18 to 36 yr., with sex and age randomly distributed in the three groups.

Procedure.—The experiment was run in three conditions: Cond. A used Inspection Fig. 2, Cond. B used Inspection Fig. 3, and Cond. C used Inspection Fig. 4. In Cond. A the left-hand dash on the test figure, i.e., the dash under the retinal location of the nonfigural half of the inspection figure, was moved while the right-hand dash was stationary. In Cond. B the reverse was true, i.e., the right-hand dash moved and the left-hand dash was stationary. In Cond. C the test figure situation was identical with the situation in Cond. A, i.e., the left-hand dash moved and the right-hand dash was stationary. Each condition was divided into two halves: the first half of each condition consisted of a determination of each *S*'s presatiation point of subjective equality (PSE) for horizontal alignment of the dashes on the test figure, *S* having first fixated Inspection Fig. 1. The second half of each condition consisted of a determination of each *S*'s postsatiation PSE for horizontal alignment of the dashes on the test figure after having fixated Inspection Fig. 2, 3, or 4, depending on the condition.

In all conditions *S* was seated in the center of a darkened 13 × 13 ft. room. Six feet in front of him, and slightly to his left, was placed Inspection Fig. 1, and at the same distance but slightly to his right, the test figure. The screens containing the figures were 2 ft. apart at their inner edges and were turned slightly toward *S* so that he could face either of them directly by slightly turning his head. Two 60-w. frosted bulbs, shielded from *S* by reflectors, were directed at the screens. The *S* was instructed that he could earn from $1.00 to $2.00 in this experiment. His task, throughout the experiment, would be to give his judgment as to whether or not the two dashes on the screen to his right were horizontally aligned. (This instruction was amplified by an analogy between the experimental task and the alignment of an image in a split image range finder.) The *E* was interested only in judgments based upon the appearance of the dashes while *S* was focused on the cross on the screen to his right. The *S* was told that the consistency and accuracy of his judgments might well be a function of the consistency with which the foregoing instruction was followed. In order to facilitate such consistency, he would be given practice in focusing before making each judgment, and this practice would take place on the screen to his left. In order to avoid the interference of afterimages, from his periods of practice focusing, upon the judging screen, he would practice his focusing with one eye and make his judgments with the other eye. (Eye order was reversed from *S* to *S*.)

The *S* was given a 3 × 5 in. white index card with which to cover the closed eye while focusing and while judging. He was cautioned against squeezing the closed eye tightly shut, in order to avoid blurring his vision when he was required to use the closed eye. The *S* was instructed that, 5 sec. before he would be asked for his judgment, *E* would say "ready," but that he was not to move until 5 sec. later when *E* would say "now."

The first fixation period for each *S* was 3 min., and subsequent fixation periods, interpolated between each judgment, were 30 sec. The *S* started fixation when *E* said "begin." At the "now" signal *S* covered, with the index card, the eye with which he had been observing the inspection figure, opened the other eye, and focused on the fixation cross of the test figure. At this point the two dashes on the test figure were covered with the black card moved into position by *E* as soon as *S* had begun his fixation of the inspection figure. As soon as *E* judged, by observation, that *S* was focused on the test figure, he lowered the black card exposing the two dashes to *S*. (The *E* estimates that in no case did more than 1 sec. lapse between the "now" signal and the lowering of the black card.) The *S* immediately gave his judgment by saying "same" if the dashes appeared to him to be aligned, or by saying "different" if they appeared misaligned. As soon as *S* gave his judgment he resumed fixation of the inspection figure. The *E* immediately returned the black card to its masking position and, in the ensuing fixation period, moved the apparatus to its next position and recorded *S*'s judgment.

Four judgments of "same" were obtained for each *S* in the first half of each condition. On all trials the method of limits was employed. The first series of judgments leading to a statement of equality by *S* started with the movable dash 2 mm. below the fixed dash. The movable dash approached objective alignment with the fixed dash by .5 mm. on each trial. The second series of trials started with the movable dash 2 mm. above objective alignment with the fixed dash. Series 3 and 4 were similarly alternated. After four judgments of equality had been obtained *S* was given 5 min. of rest. At the expiration of this 5-min. period, Inspection Fig. 1 was replaced with Inspection Fig. 2, 3, or 4, depending on the condition.

The *S* was instructed that he had already earned $1.00 by his participation in the first half of the experiment, and that he could earn an additional dollar by his performance in the second half of the experiment. It was explained to *S* that the only change in the situation would be that he would be asked to practice his focusing on a more difficult figure. He was instructed that he was to focus on the upper lip of the profile before him as if the upper lip were the center of a cross. (In the case of Cond. C, i.e., where Inspection Fig. 4 was used, *S* was instructed to focus on the cross at the center of this figure.) The *S* was told that if he could maintain the consistency of his focus on this more complex figure his judgments would

TABLE 1

SUMMARY OF *S*s' JUDGMENTS

S	Cond. A			Cond. B			Cond. C		
	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
1	+.25	+.12	−.13	−.37	−1.0	−.63	−.50	−.50	0.00
2	−.12	−.37	−.25	−.62	−.62	0.00	−.50	−.37	+.13
3	0.00	−.12	−.12	−.37	−1.0	−.63	−.75	−.50	+.25
4	0.00	−.12	−.12	−.12	−.50	−.38	−.25	−.25	0.00
5	0.00	−.12	−.12	−.12	−.37	−.25	−.62	−.50	+.12
6	−1.12	−1.25	−.13	−.37	−.87	−.50	0.00	+.25	+.25
7	0.00	0.00	0.00	0.00	−.25	−.25	−.12	−.25	−.13
8	0.00	−.25	−.25	+.12	−.25	−.37	−.12	+.25	+.37
9	0.00	0.00	0.00	−.12	−.62	−.50	+.25	−.37	−.67
10	+.37	0.00	−.37	0.00	−.25	−.25	−.25	−.75	−.50
11	+.25	−.12	−.37	0.00	−.75	−.75	−.50	−.75	−.25
12	+.37	0.00	−.37	+.12	+.25	+.13	+.50	−.25	−.75
13	+.62	+.25	−.37	−.12	−.25	−.13	−.37	−.37	0.00
14	−.25	−.50	−.25	+.25	−.25	−.50	0.00	+.37	+.37
15	+.50	−.75	−1.25	—	—	—	0.00	0.00	0.00

Note.—Each pre- and each postscore is an average of four judgments. All scores are in millimeters.

maintain or improve their level of accuracy. He was informed that if his judgments in this second half of the experiment were equal to, or better than, his judgments in the first half of the experiment, he would be paid an additional dollar.

The procedure followed was identical to that described above for obtaining the presatiation PSE. Each *S* was paid the additional dollar without reference to his performance. The total time required to run an *S* varied from 25 to 50 min. as a function of the number of trials required to obtain each judgment of equality.

All intervals were timed with a stop watch.

RESULTS

The hypothesis was tested by determining the direction of change in the PSE for each *S*, from his pre- to postsatiation judgments. Since the movable dash was always on the nonfigural side in Cond. A and B, "greater displacement downward" was always taken to mean greater satiation of the figural side. "Greater displacement upward" was always taken to mean greater satiation on the nonfigural side. In Table 1, a score preceded by a "minus" sign indicates that *S* said the two dashes were horizontally aligned when, objectively, the movable dash of the test figure was lower than the stationary dash. A positive score indicates *S* judged the dashes to be aligned when, objectively, the movable dash was higher. In the computed values in Table 1, i.e., "the difference scores," a negative value indicates that *S*'s postsatiation PSE placed the movable dash lower than it had been placed by his presatiation PSE. The converse, a positive score, indicates that *S*'s postsatiation PSE placed the movable dash higher than it had been placed by his presatiation PSE. Zero indicates that there was no difference between pre- and postjudgments. All scores given are the average of four judgments. All probabilities were determined with the sign test (two-tailed), since the direction of change was the crucial issue.

The points of subjective equality for 13 of the 15 *S*s in Cond. A changed, from pre- to postsatiation measurements, in the direction of greater displacement downward. The remaining 2 *S*s showed no change in their points of subjective equality. The probability of obtaining this finding by chance is less than .01.

The points of subjective equality for 12 of the 14 *S*s in Cond. B changed, from pre- to postsatiation measurements, in the direction of greater displacement downward. Of the remaining 2 *S*s, one showed greater displacement upward and the other showed no change in his PSE. The probability of the finding occurring by chance is less than .01.

The changes in points of subjective equality for *S*s in the control condition where the movable dash was on the left-hand side were as follows: 5 *S*s showed greater displacement downward, 6 *S*s showed greater displacement upward, and the remaining 4 *S*s showed no change in their points of subjective equality.

DISCUSSION

The results indicate that "satiation" occurred to a greater degree on the figural side. The control condition demonstrated that the method does not, itself, create an artifactual difference. It is concluded that a phenomenal nonuniformity in an objectively uniform contour will produce a differential effect in a figural aftereffect paradigm.

The direct implication of the present finding is to cast serious doubt on attempts to derive contour perception solely from the activity of peripheral mechanisms and their unmodified cortical representations. The effect reported in this paper depends upon a stimulus property which cannot be defined entirely at the objective stimulus level, i.e., definition is in part a function of an experienced organization of the stimulus. If the phenomenal status of the contour is assumed to be the variable which produced the differential effect obtained, it would seem clear that the role of central processes in modifying stimulus input must be included in an adequate description of the phenomenon. The fact that, in addition, in figures similar to that employed in the present experiment, *S*s can reverse the figure-ground relationship (Woodworth & Schlosberg, 1958, p. 404), suggests that the relevant central processes, while at present unspecifiable, are at the level of integrative function.

REFERENCES

KÖHLER, W., & WALLACH, H. Figural aftereffects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

MARSHALL, W. H., & TALBOT, S. A. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. *Biol. Symp.*, 1940, **7**, 117–164.

OSGOOD, C. E., & HEYER, A. W., JR. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (2nd ed.) New York: Holt, 1958.

(Received September 7, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 201–205

A THRESHOLD DIFFERENCE PRODUCED BY A FIGURE-GROUND DICHOTOMY[1]

BERNARD WEITZMAN

Creedmoor Institute for Psychobiologic Studies, Queens Village, New York

A previous finding (Weitzman, 1963) suggested the possibility that division of the perceptual field into figure and ground regions results in a threshold difference between figure and ground contours. To examine this possibility discrimination thresholds for 2 lines were determined, in tachistoscopic presentation, for 29 college student and student nurse Ss. Fixation was central between the lines, one of which was contained in a figural region, and the other in a ground region. Retinal equivalence of the 2 lines was indexed by an interpolated task. The prediction that the figural contour would have a lower discrimination threshold than the ground contour was supported, at the .05 level, by the results. The finding was interpreted as requiring that the possibility of modification of stimulus input, as a result of the dichotomizing of stimuli into figure and ground regions, be included in any complete description of contour perception.

In a previous paper Weitzman (1963) demonstrated that the contour of a figural region produces an aftereffect of greater magnitude than that produced by the contour of a ground region. The aftereffect of a stimulus contour is modified in that paradigm by its position in a figure-ground dichotomy. The modification, as a correlate of a phenomenal event, implies the necessity of investigating the role of central processes in contour perception. An adequate theoretical model must, therefore, include not only the modification of psychological processes by stimulus input, but the converse as well.

In attempting to explain the obtained differential effect of the previous experiment it may be hypothesized that there exists a threshold difference between the figural and the ground contour. That is, (assuming either the Köhler-Wallach, 1944, or the Osgood-Heyer, 1952, model as a frame of reference) a threshold difference seen as a functional increase in stimulus intensity for the figural contour, should result in differential intensity in the CNS processes. A threshold differential favoring the figural contour would result in a relatively more intense process in the CNS region corresponding to the figural contour. By comparison in the central region corresponding to the ground contour there would be a less intense process. This difference in intensity of central processes, it may be reasoned, would be reflected in the obtained figural aftereffect of greater magnitude on the side of the figural contour. While it remains an empirical question whether or not greater stimulus intensity will always produce a greater figural aftereffect, there is sufficient basis in the literature for using the assumption for hypothesis generation. Köhler and Wallach (1944, p. 330) present the theoretical postulate that greater intensity of CNS processes will result in larger figural aftereffects, and report a verifying experiment. The finding of Hochberg and Triebel (1955) that brightness differences are necessary for the production of figural aftereffects implies the correctness of Köhler and Wallach's speculation. The findings of Marquart (1954) and of Freeburne and Hamilton (1949) contain the same suggestions. Finding a threshold difference to obtain would not only contribute to understanding of the previous finding, but would also provide a psychophysical description of one effect of the organization of the visual field into a figure-ground dichotomy.

---

[1] This paper is based upon a portion of a dissertation submitted to the Psychology Department of the Graduate Faculty of Political and Social Sciences of the New School for Social Research, in partial fulfillment of the requirements for the PhD degree. The author wishes to express his gratitude to his committee member, Gustav Levine, for his inestimably valuable concern and criticism at every stage of the research, and in the preparation of this paper. The author also wishes to thank J. R. Whittier, Director of Psychiatric Research at the Creedmoor Institute for Psychobiologic Studies, for the patience and support which made the undertaking and completion of this study possible. The tachistoscope employed in this experiment was designed and built by the Scientific Prototype Corporation of New York City.

The present paper reports the results of an investigation designed to answer the following question: When figure and ground contours are positioned symmetrically with reference to the retina, does division of the perceptual field into figure and ground regions result in a lowering of the threshold for accurate discrimination of the figural contour, as compared with the ground contour? It is hypothesized that there is such a resulting difference in thresholds.

METHOD

Materials.—Three figures were constructed. Figure A consisted of a black outline square, 2 in. on a side, vertically bisected by an outline profile of a human face. The profile faced to the viewer's right, i.e., the face occupied the left side of the figure. An eye was drawn in an appropriate position on the face. The profile of a human face was used in this design in order to bias strongly the direction of figural organization. Pilot investigations had indicated that the use of a nonsense contour with instructions to *S* to organize one side as figural was inappropriate in a test of the present hypothesis because of the high rate of spontaneous reversals reported by *S*s. The bottom contour of the square was a straight line containing two gaps. Each gap was equidistant from one side of the square, and each was equidistant from the center of the bottom contour. The gaps were each $\frac{3}{16}$ in. long. An eye was drawn on the face further to enforce the desired direction of organization, and to lower the threshold of figure-ground dichotomizing. This was deemed necessary because pilot work had indicated that, without the eye, the particular profile employed sometimes had a higher threshold than the gaps, whereas, in order to test the hypothesis, a lower threshold was needed. Figure B was a duplicate of Fig. A with the bottom contour deleted. Figure B was introduced, both as a device to further lower the figure-ground threshold, and to increase the accuracy of scoring of correct responses. The necessity for the use of this device became apparent in the course of pilot investigations. Figure C was a rectangle, $\frac{1}{2}$ in. high and $\frac{1}{4}$ in. wide. Drawn within the borders of this rectangle were seven letters and numbers in three rows. Row No. 1 contained the number 7 and the capital letter A. The second row contained the number 1 and the letters B and X. The third row contained the number 4 and the letter W. A pre- and postexposure field was prepared consisting of a black dot on a buff field. The dot, serving as a fixation point, was so positioned that it coincided, spatially, with the geometrical center of Fig. A and C. Figures and fixation field were drawn on a buff colored mat board.

Figures A and C were presented in a three-channel, Dodge-type tachistoscope, electronically controlled. Figure A, when presented in the first field of the tachistoscope, delivered a direct view of the drawing to *S*. When presented in Field 3, Fig. A appeared as a mirror image, its profile facing to the left. In this case Fig. C was presented in Field 1, giving *S* a direct view of its contents. When Fig. A was presented in its direct view, a mirror drawing of Fig. C was presented in the third field of the tachistoscope, delivering, from this position, a noninverted image of Fig. C and its contents. The second field of the tachistoscope was always used for exposure of the fixation dot as a pre- and postexposure field. All three fields of the tachistoscope were matched in intensity with a Photovolt electronic photometer, Model

501-M, at the viewing aperture. Each field yielded an illumination of approximately .075 ft-c.

Procedure.—The *S*s were presented, tachistoscopically, with the drawing consisting of a black outline square vertically bisected by an outline profile of a human face. The bottom contour of the square contained two gaps, one to the right of the profile and the other to the left. Exposure time was gradually increased until *S*s saw both gaps. Because a test of the hypothesis required that neither gap be favored by its retinal position, a task orientation was introduced into the experimental design which, along with instructions for the experimental task, would help to maintain central fixation. Figure C, the small rectangle, appearing in the center of the field, was randomly interspersed in the presentation series. The *S*s were instructed to discriminate the contents of this rectangle. The central placement of the rectangle and its relatively small dimensions required central fixation for maximum efficiency in reading its contents. In addition, *S*s' discriminations within this figure could be taken as a gross index of the degree to which fixation actually was central.

The experiment was run in two conditions: Cond. I used Fig. A in the first field of the tachistoscope, i.e., profile seen as facing to the viewer's right. Condition II used Fig. A in the third field of the tachistoscope, i.e., with the profile seen as facing to the viewer's left. In each condition *S* was given Fig. B to inspect before looking into the tachistoscope. He was instructed that he would be shown on a screen (*E*'s way of referring to the tachistoscope) a picture exactly like the one he was examining (Fig. B) except in one respect. The picture he would be shown on the screen would have the bottom finished in some way. The *S*'s job would be to try to describe in what way the picture was finished at the bottom. The *S* was then asked to look at the screen and asked if he saw the black dot. He was told that it was very important that he keep strongly focused on the dot whenever he was looking at the screen. He was further instructed to keep his head pressed against the rubber eyepiece so that his head would not move. He was informed that the picture would be shown very quickly at first, probably too quickly for him to see it clearly, but that it would be shown again and again, more and more slowly, until he could see it clearly. He was reminded that it was very important to keep focused on the dot.

The *S* was then informed that some of the time he would be shown, on the screen, a small rectangle with letters and numbers written inside it. When this rectangle appeared his job would be to try to read the letters and numbers. He was told that the rectangle would be presented very quickly at first but that it would be shown again and again, more and more slowly, until he could read the letters and numbers. He was told that, in this case also, it was crucial that he remain focused on the dot. He was reminded that his job was to describe how the bottom of the picture was completed when the picture was shown on the screen, and to read the letters and numbers in the rectangle when that was shown on the screen. He was reminded of the importance of keeping focused on the dot and was informed that *E* would say "ready" before each presentation in order to permit him to renew his focus.

The initial exposures of Fig. A and C were always .005 sec. The exposure time was increased, independently for each figure, in .001-sec. steps to .01 sec., then increased in steps of .01 sec. to .1 sec., and then in steps of .1 sec. to 1 sec. The order of presentation of Fig. A and C was determined by a table of random numbers. After each exposure *S* was asked to describe what he had seen, until he began to offer such descriptions spontaneously. When an *S* experienced difficulty in describing what he had seen *E* offered him Fig. B and asked him to describe, with reference to this drawing, what he had seen. In pilot studies in which the mirror image of the profile was randomly interspersed in the presentation series, *S*s who had discriminated the profile in one direction were unable (within the exposure intervals employed) to discriminate the mirror image. They reported this mirror image as a nonsense figure. Since, in the present procedure, *S* described what he had seen after each exposure, *E* felt certain of the phenomenal status of his stimulus material. In no case did an *S* report a reversal of the figure once it had been discriminated. The procedure was continued until *S* reported seeing a straight line across the bottom of the picture containing two symmetrically placed gaps; one to the right of center and one to the left. At this point the figure was removed from the tachistoscope, handed to *S*, and he was asked if this was the picture he was describing.

Subjects.—Five male graduate students and 24 student nurses were volunteers for this experiment. The 5 males and 10 of the females participated in one condition, and the remaining 14 females were *S*s in the second condition.

RESULTS

The requirement of the hypothesis is that accurate discrimination of the figural side of Fig. A should take place earlier than accurate discrimination of the ground side. Discrimination was scored as accurate when *S* reported that one side contained a gap in the bottom contour, centered on that one side. Twenty-four of the 29 *S*s discriminated the figural side accurately before the nonfigural side; 12 in Cond. I and 12 in Cond. II. Of the remaining 5 *S*s, 2 discriminated the nonfigural side first, both in Cond. I. One *S* in Cond. I and 2 *S*s in Cond. II discriminated both sides simultaneously.

The probability of such a finding occurring by chance is, for each condition independently, .05 as determined by the two-tailed sign test.

Of the 29 first discriminations of the letters and numbers in Fig. C, the number 7, $\frac{1}{8}$ in. to the left and above the fixation point at a 45° angle, was discriminated first six times in Cond. I and five times in Cond. II. The letter A, $\frac{1}{8}$ in. to the right and above the fixation point at a 45° angle, was discriminated first eight times in Cond. I and eight times in Cond. II. The remaining first discriminations were the letters X and B. The letter X lay $\frac{1}{8}$ in. to the left of the fixation point and B $\frac{1}{8}$ in. to the right. Each of these letters was discriminated first by 1 *S*. In all cases the first letter discrimination took place some time after discrimination of the figure had been established but before discrimination of the gap. There is, then, by inspection, no indication of a lateral bias in visual focus, i.e., toward or away from the figural side.

DISCUSSION

The lack of any bias in the direction of the figural side in the responses given by *S*s to Fig. C gives support to the assumption that the task and instructions were sufficient to maintain centrality of fixation.

The results indicate that accurate discrimination occurs earlier within the contours of the figure than within the contours of the ground, when figure and ground contours are given equal retinal privilege. It is concluded that lowered threshold for accurate discrimination of figural contours, as compared with ground contours, is a consequence of a figure-ground dichotomy. It is not, however, possible from the present findings to state whether this modification of thresholds is best described as figural facilitation, ground inhibition, or a combination of both types of process.

One may conceive of the present paradigm as involving the examination of a single contour in two different contexts: as part of a figure, and as part of a ground. In this instance the two contexts were presented simultaneously in order to permit a comparison of their thresholds. In analyses of the nature of contour perception, typically, only one contour is presented on a homogeneous background. This one contour may be said to be equivalent to the figural side of the present paradigm. Statements which are made as to the conditions of perception, from the results of single contour paradigms, although presenting important information as to the nature of the formation of contour percepts (e.g., analyses in terms of "on-off" mechanisms), have not been able to include the variable, here shown to be relevant, of figure-ground status.

For some psychologists, notably those with a gestalt orientation, the possibility of organizing the stimulus field into figure and ground regions has been taken as the precondition for form perception as such. Since even an isolated contour may not be said to be either without form, or unrelated to a ground, the concept of figure-ground is necessary, within the gestalt frame of reference, even in the case of the "isolated" contour. The reversibility of many figure-ground relationships has been assumed, by them, to indicate that at least partial control of perceived stimulus qualities in such situations resides in the processes of the CNS. For other psychologists (e.g., Osgood & Heyer, 1952) analyses and explanations of experimental results involving contour perception have seemed possible by invoking processes, such as "on-off" firings, which are observable by presently available physiological techniques. The present finding of a direct psychophysical correlate of figure-ground dichotomizing makes clear the importance of utilizing and further exploring the concept of figure-ground in relation to contour perception. The present paradigm also makes available a psychophysical indicator (and perhaps upon further exploration, a definition) of a figure-ground dichotomy.

REFERENCES

Freeburne, C. M., & Hamilton, C. E. The effect of brightness on figural after-effect. *Amer. J. Psychol.*, 1949, **62**, 567–569.

Hochberg, J. E.; & Triebel, W. Figural after-effects with colored stimuli. *Amer. J. Psychol.*, 1955, **68**, 133–135.

Köhler, W., & Wallach, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

Marquart, D. I. The satiational theory of figural after-effects. *J. gen. Psychol.*, 1954, **51**, 83–91.

Osgood, C. E., & Heyer, A. W., Jr. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

Weitzman, B. A figural aftereffect produced by a phenomenal dichotomy in a uniform contour. *J. exp. Psychol.*, 1963, **66**, 195–200.

(Received September 7, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 206–209

EFFECT OF SPACING PRESENTATIONS ON RETENTION OF A PAIRED ASSOCIATE OVER SHORT INTERVALS[1]

LLOYD R. PETERSON, RICHARD WAMPLER, MEREDITH KIRKPATRICK, AND DOROTHY SALTZMAN

Indiana University

3 experiments varying the interval between 2 presentations of paired associates were reported. Other pairs filled the spacing and retention intervals. Pairs consisting of a word and a number as well as pairs consisting of 2 words showed a superiority at 8 sec. and longer retention intervals when presentations were spaced. Increasing the spacing interval beyond 8 sec. did not increase recall, when a 2-sec. rate of presentation was used. A 2-factor explanation of the findings was suggested, in which cue learning was the major factor influencing recall at longer retention intervals, while a 2nd factor was confounded with cue learning at short retention intervals.

An initial investigation varying the time interval between two presentations of an individual pair of associates has been reported (Peterson, Hillner, & Saltzman, 1962). Using a technique described earlier (Peterson, Saltzman, Hillner, & Land, 1962), presentations of other pairs of words and numbers filled both a spacing interval between two presentations and a retention interval from second presentation to test for recall. It was found that the retention curve after a 0-sec. spacing interval was higher at the 2- and 4-sec. retention intervals than the retention curve after an 8-sec. spacing interval. However, at the 8- and 16-sec. retention intervals the superiority was reversed.

If one assumes that the retention curve reflects retained amount of learning in a simple way, then the finding of poorer retention the longer the interval between presentations seems quite reasonable. The retention curve drops off rapidly after one presentation and hence when a second increment to associative strength is given after an interval the summed effect of the two presentations should be less than in the case where the second increment is added before forgetting of the first learning has occurred. Although such a view is supported by the findings at 2- and 4-sec. retention intervals, the superiority of the spaced presentations at longer retention intervals stands in direct contradiction to such a view. Why should an arrangement permitting partial forgetting of the first pairing result in improved retention at these longer intervals? Before attempting an explanation it seemed wise to confirm and extend the findings. The studies to be described used the same general method that was used in the earlier studies. The initial study varied length of the spacing interval, while the retention interval was held constant at a value expected to produce superiority for spaced presentations.

EXPERIMENT I

Method

Subjects.—Thirty *S*s from introductory psychology classes at Indiana University served in fulfillment of a class requirement.

Apparatus.—A memory drum was set in a 3 × 4 ft. black screen. A sliding shutter ermitted exposure of any one of multiple

---

[1] The research reported was supported by Grant G12917 from the National Science Foundation to Indiana University.

columns of material typed on tapes up to 3 ft. in length.

Procedure.—All *S*s were tested 12 times in each of six conditions. The time between spacing of two presentations of a pair of associates was either 0, 2, 4, 8, 16, or 32 sec. The retention interval following the second presentation was 16 sec. in all conditions. Stimulus members of a pair were 3- and 4-letter single-syllable familiar words. Responses were the numbers 1–15. The pairs were presented and tested in 12 blocks of exposures. Blocks were from 26 to 34 exposures long. Presentation and test of the six conditions overlapped within a given block. A stimulus word was used only in one test over the experiment, while the same numbers were randomly assigned repeatedly in each succeeding block. A given condition was begun in every serial position within a block and was approximately equally often preceded and followed by every other condition in its first presentation. Dummy pairs began each block and filled in where needed within the blocks. Dummy pairs were repeated and tested in a manner that made them indistinguishable from pairs that were scored. The *S*s were instructed to speak everything that they saw in the drum and give the appropriate number when it was missing. Guessing was encouraged. Six seconds of blank exposures occurred between blocks. The drum turned at a 2-sec. rate.

Results

Table 1 indicates that the optimum spacing intervals were 8 and 16 sec. An analysis of variance resulted in a significant spacing effect, $F\,(5,\,145) = 2.44$, $p < .05$. The 8- and 16-sec. intervals were significantly different from the 0-sec. interval at the .05 level, using the Duncan multiple-range test. Other differences were not significant.

TABLE 1

PROPORTIONS CORRECT IN EXP. I

	Spacing Interval (sec.)					
	0	2	4	8	16	32
$\bar{X}$	.39	.43	.47	.50	.51	.45
SE	.03	.03	.06	.05	.04	.04

TABLE 2

PROPORTIONS RECALLED IN EXP. II

Spacing Interval (sec.)	Retention Interval (sec.)			
	8		60	
	$\bar{X}$	*SE*	$\bar{X}$	*SE*
0	.54	.03	.33	.03
8	.61	.04	.43	.04
60	.60	.04	.43	.04

EXPERIMENT II

Method

A second experiment was planned to test the spacing intervals extended to lengths approximating the intervals between successive presentations of an item in list learning. At the same time each presentation was made to involve two exposures, as is also the case in paired-associate learning using the anticipation technique. However, the heterogeneity of events accompanying the use of the anticipation technique was considered an unnecessary complication, and both members of a pair were presented together on the first exposure of a presentation, while the stimulus was presented alone on the succeeding turn of the drum. The *S* was instructed to read aloud everything he saw in the drum aperture and in addition to say the response aloud whenever he saw the stimulus alone. Hence a double exposure was invariably a double pairing. Presentations of this kind had previously been found to be effective (Peterson, Saltzman, Hillner, & Land, 1962). Two such presentations were given, separated by either 0, 8, or 60 sec., and tested at either an 8- or 60-sec. retention interval. Sequences of exposures were organized into 18 blocks from 19 to 34 exposures long. Following every third block was a 10-sec. rest, and between the remaining blocks there were 8-sec. rests. Stimuli were similar to those in the previous experiment, and the responses were the numbers 1–19. Thirty-six *S*s from the same source as in Exp. I were tested 18 times in each of six conditions. Other details followed the procedure of Exp. I.

Results

Table 2 shows that the superiority of spaced pairings was at least as great at a 60-sec. retention interval as at 8 sec. An analysis of variance showed the spacing factor to be significant at the .01 level, F (2,70) = 5.81. Retention was also significant, F (1, 35) = 48.8.

EXPERIMENT III

Method

The experiments described previously used numbers as responses, and these were repeatedly assigned to different words over the course of the experiment. Experiment III used words as responses which were used only once, with another word as stimulus. The words were all single-syllable familiar words. There were three presentation conditions: a single presentation, two presentations with no intervening exposures, and two presentations with four intervening exposures on the drum. Test for retention was given either after one other exposure or four other exposures. Twelve *S*s were tested at a 2-sec. rate, and 12 other *S*s were tested at a 4-sec. rate. A given *S* was tested 32 times in each of six conditions. The 32 blocks were from 29 to 38 exposures in length. Eight seconds of blank exposures separated blocks, save for three changes of tape which lasted approximately 30 sec. each. Although both experiments were run in the same week by the same *E*, one was run before the other, and assignment of *S*s to the two different rates of presentation was not random.

TABLE 3

PROPORTION OF CORRECT RECALLS IN EXP. III

Retention Interval	One Presentation	Two Massed Presentations	Two Spaced Presentations
2-sec. presentation rate			
2 sec.	.79	.87	.79
8 sec.	.14	.26	.29
4-sec. presentation rate			
4 sec.	.67	.78	.75
16 sec.	.23	.26	.45

TABLE 4

ANALYSES OF VARIANCE OF PROPORTIONS CORRECT: EXP. III

Source	*df*	Presentation Rate			
		2 sec.		4 sec.	
		MS	*F*	*MS*	*F*
Presentation (P)	2	.061	10.1**	.133	13.3**
Retention (R)	1	6.216	310.8**	3.138	84.8**
*S*s	11	.022		.111	
P × R	2	.044	7.3**	.073	7.3**
P × *S*s	22	.006		.010	
R × *S*s	11	.020	3.3**	.037	3.7**
P × R × *S*s	22	.006		.010	

** $p < .01$.

Results

The mean proportions correct are shown in Table 3. The 2-sec. rate conditions will be considered first. It is apparent that two presentations were not better than one unless they were massed. The interaction between spacing interval and retention interval found in the earlier study (Peterson, Hillner, & Saltzman, 1962) was again present. At the 8-sec. retention interval a slight superiority for the spaced condition over the massed condition appeared. Both double presentation conditions were notably superior to a single presentation at the 8-sec. retention interval. An analysis of variance of the proportions correct is shown in Table 4. Both main effects as well as the interaction were significant at the .01 level.

Looking next at the 4-sec. rate, the interaction between spacing interval and retention interval was again evident (Table 3). The analysis of variance (Table 4) again showed significant main effects as well as a significant interaction.

Comparison of the 2-sec. rate with the 4-sec. rate should be made with caution, since *S*s were not randomized between the two groups. However, the differences are suggestive and relevant to the problem of the nature of short-term forgetting. Note that for the conditions where one competing presentation intervened before final presentation and test, retention dropped off in all three comparisons when the interval was lengthened to 4 sec. This occurred in spite of the longer time that the tested pair was exposed in the 4-sec. conditions. On the other hand for those single and spaced conditions where four other pairs intervened before the test, retention increased markedly when exposure time and retention interval became longer. It would seem that length of the retention interval was an important factor at the shorter intervals, while duration of the original pairing was the dominant factor at the longer intervals. It may be that events other than the interfering pairs presented in the drum are productive of forgetting for a brief time after presentation.

DISCUSSION

The experiments described in this paper together with the previous study (Peterson, Hillner, & Saltzman, 1962) indicate that a short-term retention curve does not reflect associative learning in any simple manner. It would seem necessary to postulate at least two factors to account for short-term associative retention. It may plausibly be assumed that what is being tested at intervals of 6 sec. and longer is the association between members of a pair, or cue learning. This is characterized by relatively slow retention loss and is what is ordinarily understood by the term learning. Spacing of presentations increases the effectiveness of this aspect of retention. The explanation for the superiority of spacing is not apparent. It is possible that attentional considerations are involved. Even though *S* pays attention to the degree that he is able to speak correctly what he sees on the second of two successive identical presentations, less activity on his part may result than would be the case if the presentation had not just occurred. To what extent such an explanation can be investigated experimentally remains to be seen.

It seems likely that the cue learning just mentioned is confounded at very brief retention intervals by another factor. This aspect of recall is subject to quite rapid forgetting and the decrement after spaced presentations at the shorter retention intervals would then reflect this factor. A more detailed specification of this factor awaits further research but it may be a kind of response bias or availability, or what the senior author has described as background conditioning (Peterson, 1963).

REFERENCES

PETERSON, L. R. Immediate memory: Data and theory. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963.

PETERSON, L. R., HILLNER, K., & SALTZMAN, D. Supplementary report: Time between pairings and short-term retention. *J. exp. Psychol.*, 1962, **64**, 550–551.

PETERSON, L. R., SALTZMAN, D., HILLNER K., & LAND, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

(Received September 10, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 2, 210–214

INTERACTIVE EFFECT OF STRESS AND STIMULUS GENERALIZATION ON CHILDREN'S ODDITY LEARNING[1]

LEWIS P. LIPSITT AND VINCENT M. LoLORDO

Brown University

4 groups of 14 4th-grade children performed in an oddity learning task, 2 of these receiving an easy task (colors red, green, and blue) and 2 a difficult task (red, pink, and orange). An easy and a difficult group were subjected to stress instructions involving time pressure, while the other 2 were not. The major hypothesis was that increased drive level (increased stress) would facilitate easy oddity performance but impair difficult oddity performance. This proposition was confirmed by a Stress × Difficulty interaction effect, supporting Spence's theory concerning the joint effects in complex learning of drive and interstimulus generalization.

In applying his theory to human *S*s in relatively complex learning situations, Spence (1956, pp. 199–236) has assumed that all response tendencies aroused by a stimulus are indiscriminately strengthened by the drive level (*D*) of *S*. Thus, whether increased *D* will facilitate or impair learning will depend in part on the presence and strength of response tendencies competing with the to-be-learned response. If the response designated correct is high or dominant in the response hierarchy, increased *D* should facilitate performance, while if the to-be-learned response is exceeded in strength by competing responses, impairment should result. Spence reported evidence supporting these propositions from serial and paired-associate tasks. These studies, utilizing an anxiety measure of *D* level, suggest that under conditions of high interstimulus generalization or competition, increased *D* results in performance impairment, while under conditions of low generalization or competition, increased *D* improves learning. Glucksberg (1962) has substantiated the hypothesis in a problem solving and a perceptual recognition task.

Recently, studies have appeared (e.g., Castenada, 1956; Castaneda & Lipsitt, 1959; Lipsitt, 1961) supporting the above hypotheses in children's verbal and motor learning tasks. Both test-defined anxiety and instruction-induced stress have served as indicants of *D* level. The present study is concerned with the applicability of the drive-and-competition interaction hypothesis to oddity learning in children.

The task involved the presentation of different combinations of three colored light stimuli, any two of these colors appearing on any given trial. These colors were shown in three windows, two containing the same color and the third showing an odd color. No one of the colors and no one of the three positions was differentially reinforced; the odd color was always correct. This type of task has previously proven feasible with children in a study (Lipsitt & Serunian, 1963) which demonstrated that three-position color-oddity learning improves from kindergarten to the third grade and is very easy for the latter if the stimuli are distinctive.

---

[1] This study, conducted while the second author was a senior honors candidate and National Science Foundation undergraduate fellow, was facilitated by an NSF grant (G-9681) to the senior author. The writers wish to thank Albert Ehnes, Superintendent of the Seekonk, Massachusetts Public Schools and Isabel Mullen, Principal, for their kind cooperation.

In the present study, two levels of difficulty or interstimulus generalization were introduced by manipulation of the similarity of the three colors, one group receiving three distinctive colors and the other highly similar colors. Two levels of *D* were provided through instructions designed to induce stress in half of the *S*s and not in the other.

METHOD

Subjects.—The *S*s were 56 fourth-grade children from two classrooms in one school. They were assigned successively to the four experimental groups in the order of their appearance at the experimental room, a procedure resulting in essentially random assignment of *S*s to groups. In a previous experiment all *S*s had received training on a simultaneous discrimination problem at the present apparatus, but none had previous oddity training.

Apparatus.—The apparatus is a discrimination learning unit with a remote control console. The learning unit consists of an upper panel, 26 in. wide, containing five stimulus apertures in a horizontal array, below which is a response panel of the same size sloping toward *S*. In the response panel it is possible to place up to five response buttons directly under the windows and at a distance of 7 in. from them. The number of stimulus windows and buttons used is optional, depending upon experimental requirements. Unused windows are occluded with inserts the same gray color as the panels. In the present experiment, the two end windows and the middle window at a distance of 8 in. center to center from one another were used, along with their three buttons below.

The stimuli administered on any trial are controlled from the remote console several feet to the right of *S*. The *E* determines before each trial which stimuli will be shown and which manipulandum will be correct and will activate a 200-w. white light reward (or information) panel. The reward panel is a Plexiglas window which runs horizontally the width of the apparatus between the stimulus and response panels. The stimulus lights go out immediately upon any response, which at the same time stops the clock that had been automatically activated at the moment of stimulus onset. The *E* reads from the console what the response was, whether correct or incorrect, and reads response latency from a Standard Electric clock, recording these prior to making settings for the next trial. The $1\frac{1}{2}$-in. diameter stimulus apertures are covered with milk-glass behind each of which is a cluster of bulbs, identical for each aperture. Set in front of each bulb are filters of different colors. The five colors used in the present experiment were green (Kodak Wratten No. 60), blue (No. 44), red (No. 26), orange (No. 23), and pink (No. 24).

Procedure.—When *S* was seated at the apparatus, he was informed that colored lights would be shown in the three windows, that pressing any of the buttons would turn off the stimulus lights, and pressing the correct button would in addition turn on the long white light. The *E* also demonstrated a correct response and a wrong response, using three different colors in the three windows rather than an oddity setting.

At this point the Stress and Nonstress groups received differential instructions. The two Nonstress groups were simply asked if they were ready. The *S*s in the two Stress groups were told that they had only 2 sec. in which to push one of the buttons when the lights came on. The *E* emphasized this time limit by instructing *S* to look at a stimulus array, which was turned off after 2 sec., *E* saying "That's how long you have to push the button." The *S*'s attention was called to the clock resting between *E* and *S*. Stress *S*s were then asked if they were ready, and the first oddity trial followed. On the first few trials, *E* followed correct responses with "Right that time!"

During training, *E* checked the possibility that *S* had learned to identify correctness on the basis of cues during console-dial setting. When, after learning, each *S* was asked to pick the correct button before stimulus onset, no *S* evidenced response to those cues. The *S*s were also asked at the end to name the colors shown; all *S*s could.

All 56 *S*s were run on a noncorrection procedure. Intertrial interval ranged from 10 to 15 sec., depending on time required to record the response and preset the next trial. Task difficulty or generalization was defined in terms of stimulus similarity. Half of the Stress and of the Nonstress *S*s were presented with the easy colors (red, blue, and green),

TABLE 1

Mean Number Correct Responses and *SDs* for Groups over 54 Trials

Group	Task			
	Easy (Low Generalization)		Difficult (High Generalization)	
	M	*SD*	*M*	*SD*
Stress	43.20	11.05	20.29	8.56
Nonstress	37.50	13.98	26.50	7.11

while the other half received the difficult colors (red, pink, and orange). Two levels of Difficulty and two of Stress were thus involved. In the three-position oddity task only two different stimuli are presented on each trial. Taking any one of the colors as correct, it can be paired with each of the other colors in three different positions. The order of presentation of the 18 possible settings was determined randomly and repeated three times, the only stipulation being that no one response position (left, middle, or right) would be correct on two successive trials. All *S*s received the same order of stimulus presentation. In 54 trials each setting or configuration appeared only three times.

RESULTS

The chief measure of performance is the number of correct responses over 54 trials. Means and *SDs* for this measure contained in Table 1 show that, in accord with the hypothesis, performance in the easy task tended to be facilitated by the Stress condition, whereas performance in the difficult task was impaired. Figure 1 shows the performance of the four groups over nine blocks of six trials. These data were treated in an analysis of variance (Lindquist, 1953) which permits a determination of the main effects of Difficulty, Stress, Trials, and interactions among them.

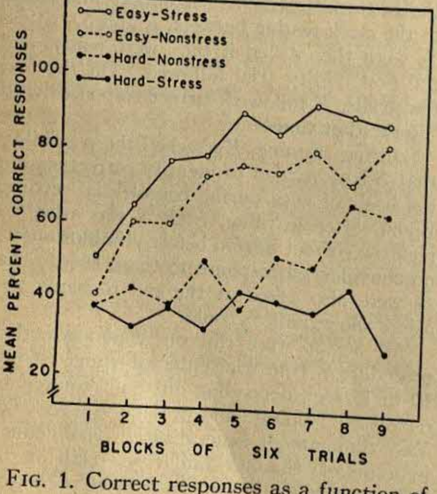


Fig. 1. Correct responses as a function of trials.

TABLE 2

Response Latencies in Hundredths of Seconds: Means and *SDs* for the Four Experimental Groups

Group	Task			
	Easy		Difficult	
	M	*SD*	*M*	*SD*
Stress	124.02	13.8	115.12	27.8
Nonstress	209.08	60.0	298.25	168.3

A reliable effect of Difficulty indicates that the similar-stimulus problem produced poorer overall performance than the distinctive-stimulus problem, $F\ (1, 52) = 33.91, p < .001$. This effect clearly establishes the presence of greater generalization in the difficult than in the easy condition and permits an evaluation of the interaction hypothesis. This interaction is represented in the Stress × Difficulty effect, $F\ (1, 52) = 4.19$, $p < .05$, based on the facilitative effect of stress on the easy task and impairment on the difficult task noted in Table 1. The main effect of Stress was not significant. Two t tests were computed to evaluate the simple effects of stress on the difficult and easy conditions, with the result that
]

the former was reliable, $t(26) = 2.01$, $p < .05$, while the latter was not ($t = 1.15$).

The main effect of Trials was significant at the .001 level, $F(8, 416) = 13.75$, indicating that for the four groups combined, learning occurred. A significant interaction between Trials and Difficulty, $F(8, 416) = 5.45$, $p < .001$, revealed that learning differed for the easy and difficult conditions, the rate being apparently greater for the easy condition. The Trials × Stress interaction, $F(8, 416) = 2.72$, $p < .01$, indicates that learning occurred differentially for the Stress and Nonstress *S*s, the Nonstress groups combined showing an apparently greater rate. Finally, the triple interaction was also reliable, $F(8, 416) = 16.10$, $p < .001$, based on a tendency for Stress to have its effect later in trials for the difficult than for the easy condition.

Reaction time data for each of the groups are in Table 2. Because the variances were heterogeneous, an extension of the median test (Siegel, 1956) was run, indicating that the differences among the medians of the groups were reliable, $\chi^2(3) = 34.84$, $p < .001$. Mann-Whitney *U* tests made on all group pairs revealed that for both the easy and difficult conditions separately, Stress produced significantly shorter latencies relative to Nonstress, reliable at the .001 level. Furthermore, while there was no reliable difference between latencies in the easy and difficult conditions for Stress, the difference between the easy and difficult for Nonstress was reliable at the .05 level. Thus it appears that Stress instructions were effective in facilitating faster responding, but under Nonstress response speed was determined by the difficulty of the task.

DISCUSSION

The results support Spence's (1956) interaction hypothesis for complex learning tasks. Increased *D*, defined here in terms of stress instructions, did have a differential effect on easy and difficult tasks. As expected, *D* facilitated performance on a task involving few competing responses (or little interstimulus generalization) but tended to impair performance on a task involving strong competing responses.

In the present oddity task, response to no specific stimulus was consistently correct; each color and spatial location was equally often incorrect as correct. This problem could be learned only by responding to the stimulus which was, on any given trial, different from other stimuli present. Thus the interaction hypothesis is relevant to learning situations other than those in which the same responses must always be paired with the same stimuli, as in paired-associate and serial-learning tasks.

Recent evidence suggests that test-defined anxiety and instruction-induced stress have similar effects upon learning performance in children. Castaneda and Lipsitt (1959) demonstrated that stress facilitates motor learning performance when the responses to be learned are already high in the hierarchy, but impairs such performance when there is competition of other responses with the to-be-learned response. The same effect (Castaneda, 1961) was produced utilizing anxiety scale scores to differentiate *D* levels. A study of verbal paired-associate learning was conducted recently (Lipsitt, 1961) in which the S-R pairs to be learned were manipulated so as to produce high generalization for some pairs and low for others. Two replications, one involving two levels of anxiety score, the other involving differential stress instructions as defining conditions for *D*, produced similar results; high anxiety and high stress impaired performance on the high competition pairs, but significant differential performance due to anxiety or stress did not appear on the low competition pairs. These experiments, like the present one and the Glucksberg (1962) study, suggest that the interaction hypothesis need not be restricted merely to those situations in which *D* is test defined.

REFERENCES

CASTANEDA, A. Effects of stress on complex learning and performance. *J. exp. Psychol.*, 1956, **52**, 9–12.

CASTANEDA, A. Supplementary report: Differential position habits and anxiety in children as determinants of performance in learning. *J. exp. Psychol.*, 1961, **61**, 257–258.

CASTANEDA, A., & LIPSITT, L. P. Relation of stress and differential position habits to performance in motor learning. *J. exp. Psychol.*, 1959, **57**, 25–50.

GLUCKSBERG, S. The influence of strength of drive on functional fixedness and perceptual recognition. *J. exp. Psychol.*, 1962, **63**, 36–41.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

LIPSITT, L. P. Effects of anxiety and stress on children's paired associates learning. Paper presented at Eastern Psychological Association, 1961.

LIPSITT, L. P., & SERUNIAN, S. A. Oddity-problem learning in young children. *Child Develpm.*, 1963, **34**, 201–206.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received September 14, 1962)

Journal of
Experimental Psychology

VOL. 66, NO. 3 SEPTEMBER 1963

DISCRIMINATION REACTION TIME FOR A 1,023-ALTERNATIVE TASK

ROBERT SEIBEL[1]

International Business Machines Research Center, Yorktown Heights, N. Y.

Stimuli were all possible patterns of 10 lights. Responses were corresponding patterns of simultaneous key depressions in a 10-key keyboard, (1 key for each finger). After more than 75,000 discrimination reaction times (DRTs) for each of 3 Ss, average DRT under the 1023-alternative condition exceeds that under a 31-alternative condition by less than 25 msec. DRT does not increase linearly with information transmitted. Learning curves for the 1023-alternative task are described by: $Y' = (A)(X + E)^B + C$. Idiosyncratic and relatively long-term performance shifts limit the precision of fit. r^2 alone was not an adequate criterion of goodness of fit.

The function relating Discrimination Reaction Time (DRT) and average information transmitted per response (approximately $\log_2$ of number of alternatives if error rate is low) has been the subject of conflicting reports in recent literature. Fitts and Switzer (1962) summarize the controversy pointing out that "One of the most generally accepted empirical laws governing perceptual-motor behavior . . ." is that DRT increases linearly with information. "A few studies, however, have revealed instances in which RT is essentially independent of information." Other investigators (Mowbray & Rhoades, 1959) conclude that DRT and information are independent. Seibel (1962) suggested that the function was one in which DRT increased from two to approximately eight alternatives, and showed trivial further increase no matter how many additional alternatives were added to the task, with the provision that adequate practice was provided and motivational factors were controlled (the effects of practice and motivation are dealt with in considerable detail).

Several models have been proposed to account for linearly increasing DRT with $\log_2$ of increasing numbers of alternatives (e.g., Hick, 1952; Luce, 1959; Rappaport, 1959; Stone,

---

[1] The author wishes to gratefully acknowledge the assistance of Jean Kirby in handling the many administrative aspects of the experiment. The computer program for accomplishing the curve fitting was written by H. W. Morrison, and modified many times to meet the needs of the present study. The author wishes to gratefully acknowledge his assistance and his many helpful suggestions.

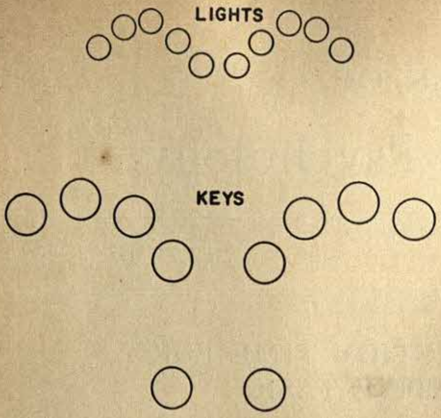


FIG. 1. Arrangements of the stimulus lights and response keys.

1960). Both Leonard (1961) and Fitts and Switzer (1962) suggest that the slope of the linear relation approaches zero under the special conditions of overlearning and/or very high compatibility. If the relation between DRT and information shifts with learning, then it should be possible to demonstrate such a shift in the laboratory, and an adequate "model" must be able to account for it.

The present experiment examines DRTs to a very large (1,023) and initially unfamiliar set of stimulus-response alternatives (though a small subset was highly familiar), traces performance till it shows little further change with additional practice, and compares DRTs in the 1,023-alternative situation with those in a 31-alternative situation. It was hypothesized that DRTs for the 1,023-alternative task would decrease with learning until these DRTs were little or no different from DRTs for a 31-alternative task, thus supporting the present author's previous findings and conclusions with respect to the relation between DRT and number of alternatives (Seibel, 1962).

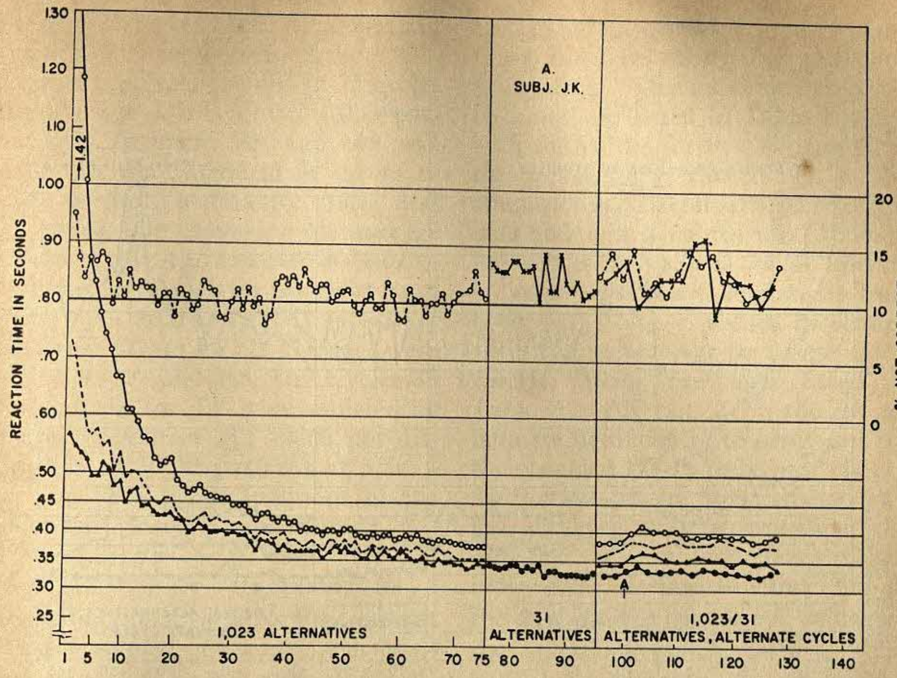
METHOD

Subjects.—Because the study called for approximately 200 experimental sessions it was not possible to use the usual hired *S*. The three *S*s in the experiment were: the author (RS), the author's assistant (JK), and another research assistant (JT).

Apparatus and procedure.—The DRT task involved an array of stimulus lights and response keys as depicted in Fig. 1. Allowing any one of the possible patterns of 10 lights (or 10 keys) yields 1,023 unique alternatives $(2^{10} - 1)$. The *S*s placed their hands over the keyboard so that a finger was positioned over each key. In response to the onset of one or more stimulus lights, *S* depressed those keys corresponding to the lights illuminated. The stimulus lights were small neon bulbs (NE-2) set in clear plastic holders in an aluminum panel. The entire stimulus array could be enclosed within a rectangle 4¼ in. long × 1⅛ in. high. The display was positioned approximately 3 ft., or more, from *S*'s eyes (except where noted in Fig. 2 for JT), thus falling within a visual angle of 7° or less. Stimuli were arranged in four different random sequences (and a fifth introduced part way through the experiment) and were automatically presented via prepunched paper tapes.

The DRT was measured from the onset of the stimulus pattern to the completion of the response, i.e., the depression of (all of) the corresponding key(s). When the response was completed the stimulus lights went out, the DRT (to the nearest .01 sec.) was automatically recorded, the paper tape advanced, and the next stimulus pattern presented. The time between completion of a correct response and the onset of a new stimulus pattern was slightly less than 2 sec. The time interval between successive presentations of new stimuli averaged just over 2 sec. once DRTs were less than .5 sec. Variation in the stimulus-to-stimulus interval (effected via equipment characteristics and a programed random sequence of delays of .5, .65, .8, and .95 sec.) reduced the rhythmical effect which a constant interval would have produced.

Three kinds of erroneous responses were possible: (*a*) extra keys depressed (non-called-for keys were scored as extra if they were depressed prior to .1 sec. after the depression of the complete called-for pattern), (*b*) one or more called-for keys not depressed within the "simultaneous" interval (defined as .1 sec. following the activation of the first key of a pattern), and (*c*) a combination of



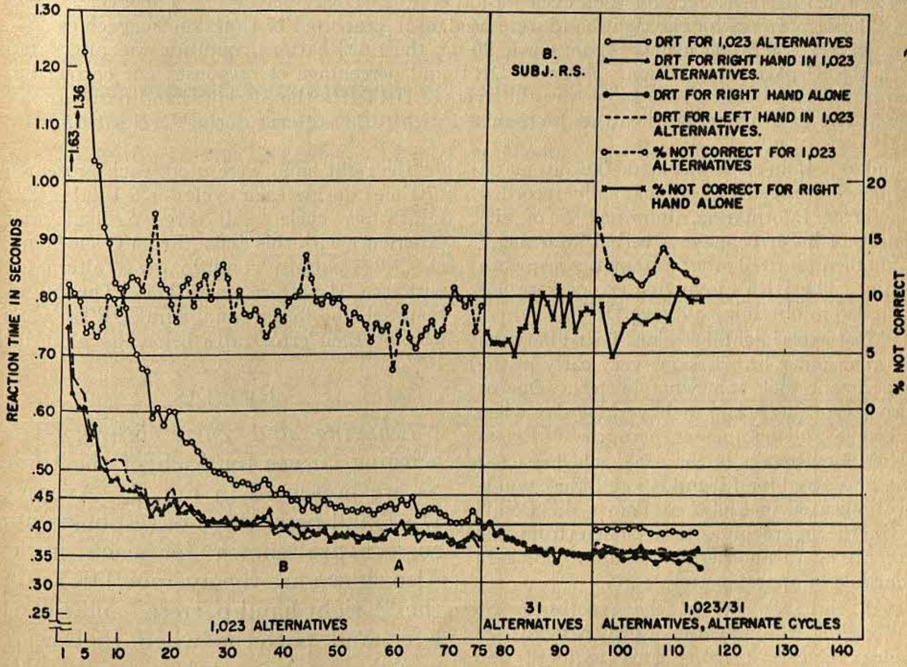


FIG. 2—Continued

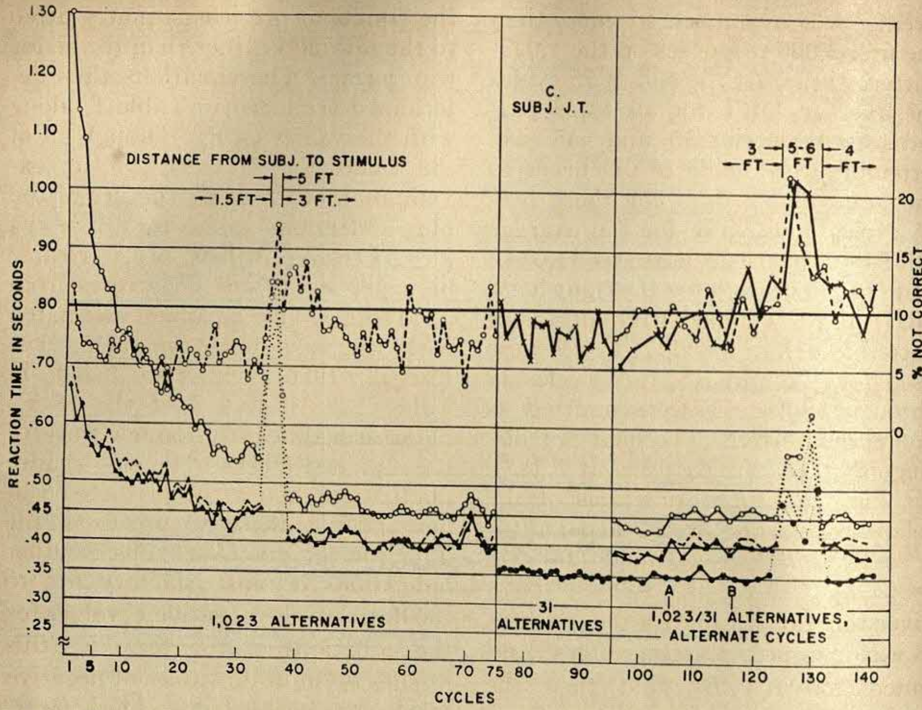


FIG. 2. Learning curves for each of the three *S*s. (For Cycles 1–75 each *S* responded to the 1,023 patterns once for each cycle—total of 1,023 DRTs per cycle. For Cycles 76–95 only the 31 patterns for the right hand were used, each presented 17 times during each cycle—total of 527 DRTs per cycle. From Cycle 96 on, the 1,023-pattern condition was alternated with the 31-pattern condition. Average DRT and percentage of responses not correct are shown for each cycle. Also shown are the average DRTs for the 31 right hand patterns and the 31 left hand patterns when they were mixed with 1,023 patterns during Cycles 1–75.)

a and *b*). If an error was made the automatic sequence stopped and *E* manually recorded the error information, informed *S* of the details of his error, gave a verbal "warning," and then started the automatic sequence again. The DRTs involving errors were not included in obtaining average DRTs.

Most experimental sessions lasted between 20 min. and ½ hr. (sessions very early in the experiment took somewhat longer). One or two sessions were run each working day when schedule and equipment permitted. Excessively long breaks in the usual schedule (1–2 mo.) occurred for JT and RS and these points are indicated in Fig. 2 at Points B. Points A in the figure indicate a change from one laboratory to another, involving a break in schedule of approximately 1 mo.

All three *S*s started the experiment by proceeding through the 1,023 patterns 75 times (Cycles 1–75). This required 150 experimental sessions. None of the *S*s had prior experience with this task. For the next 20 sessions, Cycles 76–95, only the 31 patterns for the right hand were used, each presented 17 times during each cycle for a total of 527 DRTs per cycle. All three *S*s had prior experience with this task. From Cycle 96 on, the 1,023-pattern condition was alternated with the 31-pattern condition. The *S*s attempted to obtain minimum DRTs while keeping their error rates below (or close to) 10%.

RESULTS

Learning and curve fitting.—The learning curves for each of the three *S*s are presented in Fig. 2. Average DRT and percentage of responses not correct are shown for each cycle. Also shown are the average DRTs for the 31 right hand patterns and the 31 left hand patterns when they were mixed in the total set of 1,023 patterns.

For each of the three *S*s DRTs

average well over 1 sec. for more than the first 3,000 responses on the 1,023-pattern task. At the end of 75 cycles the average DRT for all 1,023 patterns is between .35 and .45 sec. Attempts were made to fit curves to each of the *S*'s data for these first 75 cycles, one curve for the average DRTs of the 1,023 patterns (1,023), and one curve each for the right hand (RH) and left hand (LH) subsets of patterns. Data for all 75 cycles were used for JK and RS, but Cycles 35 through 38 for JT were omitted as nonrepresentative (JT could not distinguish the light patterns at 5 ft.—see Fig. 2). First, functions of the following forms were fitted: (*a*) $Y' = (A)(B)^X + C$, and (*b*) $Y' = (A)(X)^B + C$. In solving for a given function a value for the constant *B* was assumed, a set of values computed for $f(X, B)$, and then the computed values least squares fitted to the obtained ones so as to minimize $(Y' - Y)^2$, thereby obtaining the constants *A* and *C* and the correlation $r_{Y'Y}$. The procedure iterated on *B*, calculating an *r* each time, and chose the *B* leading to the largest *r*. Thus, the equations are least squares fitted to the raw data rather than to the log transforms. The equations thus determined are listed in Table 1, along with the values for r^2. Though all of the values for r^2 exceed .93, an examination of the run structure of the plus and minus values for $(Y' - Y)$ for successive values of *X* clearly indicates significant departures from good fits. The *z* values associated with the runs tests for each curve (Siegel, 1956) are also listed in Table 1. All but 2 of the 18 *z*'s indicate significantly too few runs for the obtained DRTs to be only chance fluctuations about the predicted values, and even these two are also in the direction of too few runs. Other indications of unsatisfactory fits to the data are the negative *C* values for the equations of the form *b*, (this implies asymptotic values of negative DRTs); and the large values for *C* for the equations of the form *a*, (these asymptotic values are larger than many of the DRTs actually obtained). Thus, while the fitted functions account for more than 93% of the variance of the data, the fits are

TABLE 1

CURVE FITTING FOR LEARNING CURVES

	$Y' = (A)(B)^{(X)} + C$					$Y' = (A)(X)^B + C$				
	A	*B*	*C*	r^2	*z*	*A*	*B*	*C*	r^2	*z*
1023	1.016	.865	.407	.9714	−7.54**	1.307	−.445	.170	.9915	−6.61**
JK RH	.220	.948	.345	.9788	−1.45	4.132	−.015	−3.532	.9634	−1.50
LH	.346	.922	.368	.9726	−2.41*	.812	−.161	−.056	.9799	−2.44*
1023	1.199	.891	.433	.9892	−4.77**	1.778	−.328	−.059	.9772	−8.01**
RS RH	.341	.886	.388	.9498	−4.30**	.497	−.346	.259	.9697	−3.32**
LH	.391	.893	.383	.9512	−4.24**	.583	−.358	.240	.9768	−3.32**
1023	.754	.923	.457	.9638	−5.35**	1.620	−.190	−.291	.9835	−3.94**
JT RH	.246	.941	.397	.9387	−1.97*	2.511	−.028	−1.837	.9306	−3.90**
LH	.271	.944	.399	.9404	−2.75**	2.800	−.028	−2.090	.9393	−2.15*

Note.—Simple exponential and power functions.
* $p < .05$.
** $p < .01$.

TABLE 2

CURVE FITTING INVOLVING A CONSTANT (E) TO REPRESENT LEARNING PRIOR TO THE EXPERIMENT

	$Y' = (A)(X+E)^B + C$					
	A	E	B	C	r^2	z
JK 1023	2.975	2	−.897	.314	.9974	−1.034
JK RH	9,564.04	47	−2.747	.328	.9797	−.934
JK LH	1.322	4	−.682	.289	.9858	−1.743
RS 1023	235.826	12	−2.062	.387	.9952	−2.350*
RS RH	.892	2	−.732	.338	.9738	−3.130**
RS LH	.746	1	−.563	.302	.9777	−4.166**
JT 1023	1.546	1	−.352	.091	.9848	−4.176**
JT RH	698,569	73	−4.891	.390	.9389	−1.927
JT LH	1.638	8	−.688	.317	.9418	−2.270*

* $p < .05$.
** $p < .01$.

statistically and theoretically not adequate.

Weighting each of the points by an estimate of the reciprocal of the variance (data from Seibel, 1962, Fig. 7, indicate the square of the mean as a good estimate of the variance) does not significantly improve the fits. Fifteen of the 18 runs tests are still significant.

In a further attempt to improve the fits a power function of the form $Y' = (A)(X+E)^B + C$ was fitted to the data, this time iterating on E and B. The rationale for E is that it represents learning prior to the learning represented by the data.[2] The resulting equations for the power function are listed in Table 2, along with r^2 and z values. For some of the curves six significant digits were not enough to find *the* maximum value for r^2 and an estimate within the maximum range is used. The equations for JK's data appear to fit well. The z values for the runs tests are not significantly different from zero. The minimum r^2 is .98. And, the values

TABLE 3

CURVE FITTING FOR RS AND JT DATA PRIOR TO INTERRUPTIONS

	$Y' = (A)(X+E)^B + C$					
	A	E	B	C	r^2	z
RS 1023	12.556	6	−1.168	.299	.9950	.201
RS RH	12.651	7	−1.738	.388	.9729	−.808
RS LH	.670	0	−.272	.142	.9783	−1.782
JT 1023	1.991	0	−.140	−.680	.9864	−1.378
JT RH	2.129	1	−.039	−1.416	.9246	−.329
JT LH	1.133	1	−.090	−.373	.9385	−.348

Note.—Cycles 1 through 39 for RS, and 1 through 34 for JT.

---

[2] A similar argument for the exponential function yields $Y' = (A)(B)^{(X+E)} + C$, but this reduces to $Y' = (K)(B)^X + C$, where $K = (A)(B)^E$, and $r_{Y'Y}$ is independent of E.

of E are consistent with its interpretation, namely, the curve for the 1,023 alternatives indicates the least experience prior to the experiment, the curve for the left hand patterns indicates a slight bit more, and the curve for the right hand patterns a great deal more. The asymptotic values corresponding to *the* maximum r^2s are reasonable as a group; however, they do not appear to be in a reasonable order for the three different curves. The value for the 1,023 curve is less than that for the right hand curve, and the value for the left hand curve is less than that for either (JK, and the other two *S*s, are "right-handed").

The equations for the RS and JT data indicate additional difficulties. The asymptotic (C) value for JT's 1,023 curve is unreasonably small, the values for E for RS's curves are inconsistent with the interpretation for E, and the runs tests indicate significantly too few runs for five of the six curves. Since JK was the only *S* to complete the first 75 cycles without a major interruption in schedule or experimental conditions, fits were made to those portions of the RS and JT data which preceded the interruptions (for RS: Cycles 1 through 39; for JT: Cycles 1 through 34). The resulting equations are listed in Table 3, along with r^2 and z values. The equations have E values which are not inconsistent with the interpretation for E, but the values for the right hand curves give little or no indication of the extensive prior practice for this subset of patterns. The asymptotic values for the RS data order themselves in the same "unreasonable" way as they did for the JK data, and the asymptotic values for the JT data are negative and therefore entirely unreasonable. However, the r^2s indicate good least squares fits (the worst fit accounts for 92% of the variance), and the z values are not significantly different from zero. Esthetically, plots of the obtained equations seem to "fit" the obtained data. Questions must be raised about the data (or the form of the equations) in order to reconcile the "unreasonable" values which the equations contain (see Discussion and Conclusions section).

Information transmitted in the 1,023-alternative task.—A lower bound to the information transmitted from stimuli to responses may be obtained for given proportions of correct responses (p) as follows: Assume equal error rates for each of the C different stimuli, and for each stimulus the proportion of error equally spread among all responses. All stimuli are equally likely. This leads to a $C \times C$ matrix in which all the major diagonal cells have the same probabilities, namely, p/C, all other cells contain probabilities $(1 - p)/(C)(C - 1)$, and row and column sums equal $1/C$. Clearly this is not what occurs with DRT data since certain stimuli are much more likely to produce incorrect responses than others, and certain incorrect responses are much more likely than others, but this model leads to the minimum possible transmission of information (T) for given p and C values. Under these conditions the formula for T is

$$\begin{aligned} T = \log C & \\ & - (1 - p) \log (C - 1) \\ & + p \log p \\ & + (1 - p) \log (1 - p). \quad [1] \end{aligned}$$

For $C = 1{,}023$ and $p = .90$, $T = 8.53$. Dividing 8.53 by .40 sec. (the average DRT toward the end of the 75 cycles) leads to an information transmission rate of 21.3 bits per sec. This value is a minimum for a 10% error rate.

The maximum is just under 25 bits per sec. (10/.40 sec.). The value for the present data must lie in between.

Comparison of 31- and 1,023-alternative tasks.—The DRTs for the 31 right hand patterns were averaged for those cycles (62–75) during which they were randomly mixed with the 992 other patterns, and for those cycles (77–94) during which only these patterns were called for. The DRTs in the 1,023-alternative condition were longer than those in the 31-alternative condition for all three *S*s, the overall average difference being .027 sec. The DRTs (in sec.) were: JK, .346 vs. .334; RS, .378 vs. .363; and JT, .406 vs. .351. For each of the 31 patterns involved, for each *S*, at least 30 differed in the same direction. Thus, the direction of difference, for all three *S*s, is statistically significant ($p < .01$). Essentially the same comparison was made for those cycles during which the 1,023-alternative condition alternated with the 31-alternative condition (Cycles 96–127 for JK, 96–115 for RS, and 96–105 for JT). For all three *S*s DRTs in the 1,023-alternative condition were longer than in the 31-alternative condition, the overall average difference being .022 sec. The DRTs (in sec.) were: JK, .353 vs. .333; RS, .351 vs. .341; and JT, .381 vs. .345. However, the difference for JK is inflated by the increase in DRTs in the 1,023-alternative condition following the change in laboratories (Point A, Fig. 2), and the difference for RS is apparently too small because of the higher error rate for the 1,023-alternative condition. The differences for each of the 31 patterns were all in the same direction for JK and JT, and 27 out of 31 were in the same direction for RS. The direction of difference, for all three *S*s, is statistically significant ($p < .01$).

Thus, examining the DRTs for the 31 patterns common to both the 31- and the 1,023-alternative conditions suggests that the 992 additional alternatives add between .02 and .03 sec. to the average of the DRTs. However, this difference cannot be attributed solely to the effects of the larger number of alternatives. For all three *S*s the average DRTs for the 31 patterns was less than the averages for the 1,023 (including the 31), the differences ranging from .037 to .049 sec. Imbedding a relatively easy (shorter DRTs) set of patterns in a more difficult set tends to increase the DRTs to the easier set (Seibel, 1960, 1962), thus accounting for at least part of the .02–.03 sec. difference obtained. Seibel (1962) indicates that under these conditions one should also get more variability in the DRTs for the isolated easier set of patterns that one gets for these patterns when they are mixed with more difficult ones. Such is the case for all three *S*s when collapsing across Cycles 96–127 for JK, 96–115 for RS, and 95–105 for JT; though the instability of the variance estimates due to "atypical" DRTs precludes any precise comparisons.

After extensive practice (61 or more cycles) the variable with the major effect on DRTs is the particular pattern. For example, of the 1,023 patterns, the DRTs for the 32 most difficult range from .08 to .11 sec. longer than the DRTs for the 32 easiest.

DISCUSSION AND CONCLUSIONS

The data of the present experiment indicate that *less* than 20–30 msec. are added to DRTs when information per response is increased from between 4 and 5 bits to between 8.5 and 10 bits. Prior data (Seibel, 1960, 1962) indicate approximately 40 msec. are added to DRTs when information per response is increased from slightly less than 1 bit to between 4 and 5 bits. Thus, while *the* function relating DRT and information per response cannot be specified by the data, the data do indicate that the function is *not* a linear increasing one, *nor* is it a straight line with zero slope. The data are, however, consistent with a function which increases from 1 to approximately 3 bits (2 to 8 alternatives), and shows little further increase from 3 to almost 10 bits (8 to almost 1,024 alternatives).

There are two aspects of the present experiment, however, which appear to be crucial to any generalizations. They are(*a*) the degree of compatibility of the stimulus-response (S-R) arrangement, and (*b*) the "uniqueness" ("particular stimulus and response codes," Fitts & Switzer, 1962) of it.

Compatibility.—The S-R arrangements of the type used in the present experiment have usually been considered highly compatible, however, Leonard (1961) argues otherwise and uses the *lack* of compatibility argument to explain why extensive practice was required before the S of the Mowbray and Rhoades (1959) study (using a similar S-R arrangement) demonstrated equivalent DRTs for 2- and 4-choice tasks. With the more compatible task of responding by pushing with the finger which was being directly stimulated with mechanical vibration, Leonard's (1959) *S*s yielded essentially equivalent DRTs for 2-, 4-, and 8-choice tasks with very little practice. Such an explanation fails, however, to account for the present author's results (Seibel, 1962) in an experiment using an S-R arrangement very similar to that used by Mowbray and Rhoades. The results indicate no differences in DRTs for 10-, 15-, 25-, and 30-choice tasks, even with relatively little practice. The concept of compatibility does not seem adequate to explain the findings, yet the degree of compatibility of the S-R arrangement cannot be ignored when attempting to generalize the findings of the present study. If, e.g., direct tactual stimulations were used for stimuli instead of the array of lights, it is quite possible that DRTs would not be found to be independent of number of alternatives (as Leonard, 1959, found for a one-key-at-a-time task) since this task requires *patterns* of key presses as responses.

Uniqueness.—With the exception of previous work by the present author (Seibel, 1960, 1962) DRT studies involving key pressing responses have limited the responses to the pressing of one key at a time. Only "continuous response" studies have utilized patterns of key presses as responses (Klemmer & Muller, 1953; Quastler, 1955). Only the key pressing studies utilizing the large numbers of unique S-R pairs made possible by the patterns, as opposed to single keys, have reported high rates of information transmission. Quastler (1955) reports transmission rates approaching 25 bits/sec for pianists playing random notes. Transmission rates exceed 21 bits/sec in the present study. The highest transmission rate reported for the one-key-at-a-time studies is that reported by Leonard (1959) for key pressing to direct tactual stimulation. The rate is approximately 14 bits/sec. Perhaps the *patterns* of lights and keys are "unique," and generalization to one-at-a-time situations is not warranted. The pattern situation might be rationalized as follows: DRT is a linear increasing function of $\log_2$ of the number of S-R channels, and each light and key combination represents a single channel. Thus DRT should increase as the number of channels (C) increases from 1 to 10, and the ($2^C - 1$) different patterns really only represent 10 different channels. This leads to the prediction that a 10-choice, *one-key-at-a-time* task should yield the same average DRT as a 1,023-choice *pattern* task—if, somehow, the effects of familiarity, motor difficulty, etc. could be subtracted out of the pattern task. Inherent in such a formulation, however, is the problem of defining what constitutes an S-R channel for, e.g., verbal responses; and how does one decide when to count numbers of alternative choices and when to count numbers of channels? The argument carries back to the ideas of coding and chunking (Miller, 1956) and seems to suggest that it is really the "subjective bit" that is determining the behavior. But, there is at present no way of specifying the subjective bit without getting into a theoretical circle. At this point in experimental time the present author prefers to count the number of alternative choices regardless of the situation, i.e., pattern, one-at-a-time, verbal, etc., until the count approaches Miller's (1956) "Magical Number Seven, Plus or Minus Two:" and from that number on consider the *limits* of performance independent of the number of additional choices.

The concept of the *limits* of performance, however, implies the opportunity for adequate practice so that additional practice produces little or no further change in performance. The results of the present study indicate that for unfamiliar and complex tasks adequate practice can mean 50,000 responses and more. One may argue that one intends only to construct a model for performance during the early phases of this learning and the limits of performance are of no interest. Any complete model must, however, at least reflect changes with practice, and must take account of the relative stage of practice in order to compare any given tasks. Specifying equivalent stages of practice for two or more complex tasks poses some extremely difficult theoretical and operational problems (Seibel, 1962), and it is proposed that consideration of the limits of performance is at present, the only way out. For the limits of performance a "direct association mechanism" seems to describe performance on tasks involving large numbers of alternatives, and it is only when the number of alternatives is less than "seven, plus or minus two" that the alternatives can be handled as a set isolated from the thousands of other S-R links of everyday life, with a consequent reduction in average DRT.

In attempting to fit equations to the data of the first 75 cycles of the present experiment many statistical and theoretical questions are raised. Grant (1962) and Binder (1963) have recently summarized many of the statistical strategy questions involved. In the present report the maximum $r_{Y'Y}$ resulting from the least squares fitting procedure was utilized as one of the criteria of goodness of fit, but the absolute values obtained, as pointed out by Grant, are "completely dependent upon the particular values of the independent variable . . ." which are involved. The introduction of the E parameter in the equation $Y' = (A)(X + E)^B + C$ further confounds this dependence. In addition, the value of $r_{Y'Y}$ (actually r^2) was found to be relatively *in*sensitive to major differences in parameter values and even to the forms of the fitted functions. For these reasons other criteria of goodness of fit were also used.

One of the other criteria used was based on an examination of the sequences of actual deviations of predicted from observed values by means of the runs test. In almost all cases the number of runs was less than would be expected if the deviations were independent, and just chance fluctuations about zero. It is true that the numbers of runs were *not significantly* fewer than chance would predict for the final equations, but they were still almost all in the direction of too few, rather than an approximately equal mixture of too few and too many. This strongly suggests that deviation values for successive values of X are not independent. The observed values tend to occur in relatively long runs on one side or the other of the predicted curve. If these runs tended to occur with any consistency in one place or another for the various curves it would be reasonable to conclude that the function was just not right for the data. But, this is not the case. The deviations apparently reflect relatively long-term shifts in concentration on error vs. speed, shifts in motivation, and other idiosyncratic factors peculiar to individual *S*s in long experiments (and perhaps short ones, too). Thus, optimizing the fit on any statistical grounds, beyond accounting

TABLE 4

EQUATIONS YIELDING MAXIMUM r^2 AND RANGES OF EQUATIONS YIELDING $r^2 = (r_{\max})^2 - .0001$, FOR JK'S DATA

	A	E	B	C	r^2	z
JK 1023	(1.891)	(1)	(−.691)	(.275)	(.9965)	−3.815**
	2.894	2	−.874	.308	.9973	−2.647**
	2.975	2	−.897	.314	.9974	−1.034
	3.052	2	−.918	.320	.9973	−1.975*
	(4.822)	(3)	(−1.084)	(.335)	(.9971)	−2.440*
JK RH	157.184	31	−1.870	.319	.9796	−1.448
	6,743.70	47	−2.656	.325	.9796	−.545
	9,564.04	47	−2.747	.328	.9797	−.934
	13,467.6	47	−2.836	.330	.9796	−.934
	703,896	81	−4.576	.335	.9796	−.545
JK LH	(1.023)	(3)	(−.570)	(.270)	(.9856)	−1.713
	1.276	4	−.656	.283	.9857	−1.966*
	1.322	4	−.682	.289	.9858	−1.743
	1.371	4	−.708	.295	.9857	−1.743
	1.745	5	−.788	.302	.9857	−1.733

Note.—Values for r^2 less than $(r_{\max})^2 - .0001$ are enclosed in parentheses.
* $p < .05$.
** $p < .01$.

for a major percentage of the variance, seems to call for the highly unusual task of fitting to a nonrandom and unknown error distribution which is different for each *S*. And, putting many "random" *S*s together as a way out leads to a host of other problems having to do with the form of a composite curve, the components of which have different parameter values and/or different forms.

Other criteria used for the curve fitting were the ones of reasonable values for the parameters, and internal consistency for these values in terms of their theoretical interpretation. In attempting to satisfy these criteria one encounters questions relative to the precisions of the estimates for each of the various nonorthogonal parameters. In order to illustrate the *lack* of precision the "best fitting" equations for JK's data are examined in more detail. Table 4 lists the equations for which r^2 is within .0001 of $r_{\max}^2$ (or as close to this tolerance as permitted by the integer iterations on E). The ranges of some of the parameter values are very large, particularly for the Right Hand (RH) data. When these difficulties are coupled with those associated with the run structure (nonindependent deviations), one is led to reconsider just what may be concluded from attempts at curve fitting.

Clearly none of the functions met all of the goodness of fit criteria set for them. Of the functions explored, the "weight of evidence" favors the function $Y' = (A)(X + E)^B + C$, and it provides a reasonably good description of the data. The experimental technique contains far too much error to arrive at a more precise description.

REFERENCES

BINDER, A. Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.*, 1963, **70,** 107–115.

FITTS, P. M., & SWITZER, G. Cognitive aspects of information processing: I. The familiarity of S-R sets and subsets. *J. exp. Psychol.*, 1962, **63,** 321–329.

GRANT, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.*, 1962, **69,** 54–61.

HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4,** 11–26.

KLEMMER, E. T., & MULLER, P. F. The rate of handling information: Key pressing responses to light patterns. *USAF HFORL memo Rep.*, March 1953, No. 34.

LEONARD, J. A. Tactual choice reactions. I. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

LEONARD, J. A. Choice reaction time experiments and information theory. In Colin Cherry (Ed.), *Fourth London symposium on information theory.* Washington, D. C.: Butterworth, 1961.

LUCE, R. D. Response latencies and probabilities. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences.* California: Stanford Univer. Press, 1959.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

MOWBRAY, G. H., & RHOADES, M. U. On the reduction of choice-reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

QUASTLER, H. (Ed.) *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955. Pp. 341–349.

RAPPAPORT, A. A study of disjunctive reaction times. *Behav. Sci.*, 1959, **4**, 299–315.

SEIBEL, R. Some factors influencing choice reaction time. Paper read at American Psychological Association meetings, Chicago, September 1960.

SEIBEL, R. Discrimination reaction time as a function of the number of stimulus-response pairs and the self-pacing adjustment of the subject. *Psychol. Monogr.*, 1962, **76**(42, Whole No. 561).

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

STONE, N. Models of choice-reaction time. *Psychometrika*, 1960, **25**, 251–260.

(Received October 10, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 227–234

RECALL OF CATEGORIZED WORD LISTS[1]

BURTON H. COHEN

Lafayette College

According to Miller's chunk hypothesis, the number of word categories (chunks) represented in the immediate word recall of a list of 20 categories of words should be the same as the number of words recalled from a list of 20 unrelated words. Using a design involving list lengths of 10, 15, and 20 chunks and 3 types of word lists (2 categorized and 1 unrelated), the hypothesis was supported—but only if the lists were matched for total presentation time. An index of interitem associative strength, derived from word association norms, showed significant correlations with within-category word recall but was not related to chunk recall.

In an earlier study of free recall of categorized word lists (Cohen, 1963), a distinction was made between two classes or categories of words. An exhaustive (E) category was defined as one where three or four words represented all or nearly all of the words of the category, e.g., FRESHMAN, SOPHOMORE, JUNIOR, and SENIOR. A nonexhaustive (NE) category was defined as one where three or four words were selected from a large number of words which nominally belong to the category, e.g., BEAR, LION, HORSE, and DOG. The *S*s were presented with lists comprising 20 categories of words. The results indicated that: (*a*) *S*s recalled words from approximately 12 to 14 of the word categories; (*b*) both E and NE word categories were equally represented by the recalled words but significantly more words of E categories were recalled; and (*c*) an index of interitem associative strength, derived from word association norms, effectively discriminated between the E and NE categories and was highly correlated with within-category word recall.

These results were interpreted as providing some support for Miller's (1956) views regarding verbal recoding. According to Miller, groups of words that go together (i.e., are related in some manner) are recoded and given a new name or label. In this fashion, a group of related words can function as a single unit in memory. If Miller's assumption is correct, a list comprising 20 categories of words and a list of 20 unrelated words are equivalent in the sense that each contains 20 units or *chunks* of information. In the former case, each unit or chunk comprises the set of related words belonging to a given category while in the latter case the unit or chunk is the set of letters comprising a given word. According to Miller's chunk hypothesis, there should be no significant difference between the number of words recalled from a list of 20 unrelated words compared to the number of categories represented by the recalled words of a list of 20 word categories.

Two experiments were conducted to evaluate this hypothesis. In Exp. I, nine groups of *S*s were presented with lists of words for free recall. Each of three groups of *S*s received lists of 10, 15, or 20 unrelated words; 10, 15, or 20 NE word categories; or 10, 15, or 20 E word categories. The words of all lists were presented at a constant rate. In Exp. II, the rate of presentation of the words of the 20-chunk lists was varied so that the relation between chunk recall and total presentation time could be assessed. In addition to the free recall data, word associations to all the words of the E and NE categories were collected.

---

[1] This research was supported, in part, by Research Grant M-4430(A) from the Public Health Service, National Institutes of Health and by the National Science Foundation's summer program of Research Participation for College Teachers held at the Pennsylvania State University.

EXPERIMENT I

Method

Construction of lists.—The 20 E categories were selected first. Though the number of E categories is quite limited, an attempt was made to satisfy the following criteria: (*a*) minimal intercategory associations, both at the level of the individual words and the category labels; and (*b*) maximum degree of exhaustiveness, i.e., a small category size. Half the categories comprised three words and half comprised four words. For use as a matching device in the selection of the categories and words of the other lists, the geometric mean of the Thorndike-Lorge (1944) frequencies of the words of each E category was computed.

Once the E categories were selected, 10, three-word and 10, four-word NE categories were selected on the basis of the following criteria: (*a*) minimal intercategory associations, both at the level of the individual words and the category labels; (*b*) minimum degree of exhaustiveness, i.e., a large category size; and (*c*) each NE category was matched with an E category according to the geometric mean of the Thorndike-Lorge frequencies of the words of the category. The NE categories were selected from a set of norms collected by Cohen, Bousfield, and Whitmarsh (1957).

Finally, a list of 20 unrelated words was compiled on the basis of the following criteria: (*a*) minimal interword associations;[2] (*b*) minimal duplication of initial letters; and (*c*) each word was matched for Thorndike-Lorge frequency with the geometric mean of one of the E categories. The words of the 20 E and 20 NE categories as well as the 20 unrelated words are shown in Table 1.

From the basic list of 20 E categories (E20), a 15-category list (E15) was formed by randomly selecting 15 of the original 20 categories with the restriction that 8 categories contain 4 words and 7 categories contain 3 words—a total of 53 words. The 10-category list (E10) comprised the random selection of 10 of the original 20 categories, 5 of 4 words and 5 of 3 words each—a total of 35 words.

From the basic list of 20 NE categories, the 15-category list (NE15) was formed by selecting the 15 NE categories which had been matched on Thorndike-Lorge frequency with the corresponding E categories. In a similar fashion, the 10-category (NE10) list was formed. The same procedure was followed for the list of 20 unrelated words (U20) to form the 15-word (U15) and 10-word (U10) lists.

Procedure.—The words of each list were typed in capital letters on adding machine tape and presented to individual *S*s at a 3-sec. rate by means of a Gerbrands memory drum, Model M1-A. Within each condition, every *S* received a different random order of the words of the list. The words were randomized such that no word of a given category was ever followed by a word of the same category. The *S*s were told that after the last word was presented they were to write the words of the list they could recall, in any order that occurred to them. Blank sheets of paper were provided for the written recall. The *S*s were informed they would have 5 min. for recall.

Subjects.—The *S*s were 135 male undergraduate students attending the summer sessions at Lafayette College and the University of Connecticut. The random assignment of *S*s, in blocks of 9, to each of the nine different lists, provided a total of 15 *S*s per condition.

Word association norms.—Word associations to all the words of the 20 E and 20 NE categories were collected. The technique for collecting the norms is described in detail in the previous study (Cohen, 1963). Briefly, by means of individual, manually-operated memory drums, *S*s exposed one word at a time and wrote their associations to one word before exposing the next word. This technique duplicated, in almost every detail, the manner in which the lists were presented to *S*s for free recall. Three different random orders of the lists were used. A total of 51 male undergraduate Lafayette College students served as *S*s for the words of each list.

---

[2] The interitem associations among the unrelated words were assessed by means of an unpublished set of norms kindly provided by James Deese.

TABLE 1

STIMULUS WORDS

Exhaustive Words				
NORTH SOUTH EAST WEST	SPRING SUMMER WINTER AUTUMN	PENNY NICKEL DIME QUARTER	MOTHER FATHER BROTHER SISTER	COCAIN HEROIN MORPHINE OPIUM
ACE KING QUEEN JACK	ADDITION SUBTRACTION MULTIPLICATION DIVISION	FRESHMAN SOPHOMORE JUNIOR SENIOR	BASS TENOR ALTO SOPRANO	HURRICANE TORNADO TYPHOON MONSOON
BLONDE BRUNETTE REDHEAD	KNIFE FORK SPOON	CAUCASIAN ORIENTAL NEGRO	LENGTH WIDTH HEIGHT	CIGARETTE CIGAR PIPE
SOLDIER SAILOR MARINE	MASCULINE FEMININE NEUTER	PROTESTANT CATHOLIC JEW	VOLT AMPERE OHM	ICE WATER STEAM
Nonexhaustive Words				
ARM LEG HEAD HAND	MOUNTAIN HILL VALLEY CLIFF	BEAN LETTUCE CARROT POTATO	ENGLAND FRANCE GERMANY RUSSIA	SAXOPHONE CLARINET OBOE TRUMPET
DIVAN DESK TABLE BED	GUN CARBINE SABER KNIFE	YACHT YAWL SCHOONER CANOE	OPAL SAPPHIRE ONYX TOPAZ	TWEED CORDUROY GABARDINE DENIM
HYDROGEN SODIUM HELIUM	BEE ANT SPIDER	GOLF HOCKEY TENNIS	RAIN FOG SNOW	WILLOW ELM OAK
HOUSE HOTEL MANSION	ROBBERY THEFT LARCENY	REVEREND PRIEST MONK	CHOLERA LUKEMIA EPILEPSY	MINUTE WEEK YEAR
Unrelated Words				
HAYSTACK BALLET COCOON MONGREL	REFEREE GARAGE JUICE FLAG	VEAL SOFA ARROW DOLLAR	TUTOR INK NOON EYE	GLUE CREEK LIBRARY WINTER

None of these *S*s had participated in the collection of the free recall data.

Results

Scoring.—The recall protocols were scored for the number of words recalled and, for the E and NE conditions, the number of categories (chunks) represented by the recalled words. In scoring word recall, credit was given only for those words which were members of the stimulus list. The *S*s were given credit for having represented a category in recall if at least one word of the category was recalled. Since the data from the Lafayette *S*s did not differ significantly from the Connecticut *S*s, both

TABLE 2

MEANS AND *SD*s OF WORD RECALL AND CHUNK RECALL

Cond.	Word Recall		Chunk Recall	
	M	*SD*	*M*	*SD*
U10	7.47	1.41	7.47	1.41
U15	9.07	1.67	9.07	1.67
U20	9.53	2.32	9.53	2.32
NE10	16.67	5.26	7.40	1.35
NE15	23.60	6.06	9.93	2.25
NE20	26.67	5.63	12.33	2.23
E10	26.27	4.37	8.67	1.23
E15	33.33	6.61	11.13	1.25
E20	38.47	4.58	12.80	1.42

Note.—Chunk recall and word recall are identical for the U conditions.

sets of data were pooled. The means and *SD*s for word recall and chunk recall for each of the nine conditions are shown in Table 2.

Word recall.—Because of the difference in list length (measured in words), a separate analysis of variance was computed for the U conditions. The results showed a significant increase in word recall as a function of increasing list length—F (2, 42) = 5.20, $p < .01$. The 2 × 3 analysis of variance for the E and NE conditions showed: a significant increase in word recall as a function of increasing list length—F (2, 84) = 80.88, $p < .001$; and a significantly greater word recall for the E conditions—F (1, 84) = 31.55, $p < .001$. The interaction term was not significant ($F < 1.0$).

Chunk recall.—A 3 × 3 analysis of variance of the chunk-recall means showed: a significant overall increase in chunk recall as a function of increasing list length—F (2, 126) = 52.00, $p < .001$; a significant difference among the three types of lists—F (2, 126) = 17.76, $p < .001$; and a

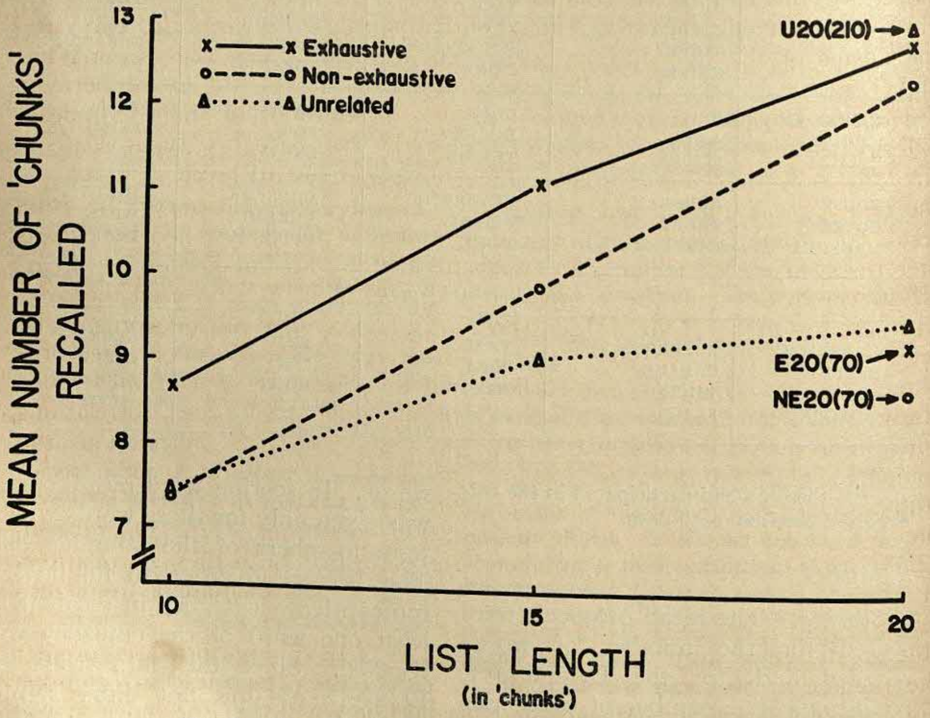


FIG. 1. Mean chunk recall as a function of list length.

significant interaction between list type and list length—$F(4, 126) = 2.83$, $p < .05$.

Inspection of the chunk-recall means, which are shown plotted in Fig. 1, suggested that the interaction was due to the divergence of the E and NE conditions from the U conditions. The following results confirmed this suggestion. (*a*) Both the E and NE conditions showed a significant linear trend with increasing list length ($p < .001$). The linear trend of the U conditions was not significant. None of the curvilinear components of variance were significant. (*b*) The means for the E20 and NE20 conditions did not differ significantly from one another while both differed significantly from the U20 condition ($p < .01$). (*c*) There were no significant differences among the U10, NE10, and E10 means. (*d*) The means for the E and NE conditions did not differ significantly, though the means of the E conditions were higher in every case than the corresponding NE means.

Evidence from two sources may be cited to show that the divergence of the chunk-recall means was not the result of some peculiarity of the present study. (*a*) The E20 and NE20 conditions may be compared with the results of Exp. III of the previous study (Cohen, 1963). Combining the two lists of the earlier study, the mean number of E categories represented in recall was 12.91 compared to 12.80 for the present study. The corresponding means for the NE categories were 12.30 and 12.33. (*b*) On the basis of a number of studies of free recall of lists of unrelated words, Murdock (1960) was able to derive a simple formula for the prediction of mean word recall. For lists of 10, 15, and 20 unrelated words presented at a 3-sec. rate, Murdock's formula predicts the following mean recalls: 7.90, 8.80, and 9.70. The agreement between these predicted values and the actual means for the U conditions is rather good. Thus, it appeared reasonable to conclude that the divergence shown in Fig. 1 was a stable phenomenon.

Relations with word association data.—In an attempt to isolate a variable which could account for the divergence, attention was directed to the word association norms. The index of interitem associative strength (IAS), as described by Deese (1959), was derived for each of the E and NE categories. This index is a measure of the degree to which the words of each category elicit each other as word associates. The mean IAS index was 23.3 (mdn. = 25.7) for the E categories and 6.2 (mdn. = 4.3) for the NE categories. As in the previous study, the mean IAS index clearly distinguished between the two category types. Using the Mann-Whitney *U* test, the difference between the distributions of the IAS indexes for the two category types was significant at the .01 level.

For a given category, the total number of words recalled by *S*s was divided by the maximum recall attainable if all *S*s who used the category had recalled all the words of the category. The ratio so obtained provided a measure of within-category recall. The rank-order correlations between the IAS index and the measure of within-category recall for the 20 categories of the E20 and NE20 conditions were .59 and .50, respectively. Both these correlations are significantly different from zero ($p < .01$).

The IAS index did not predict which of the categories (i.e., chunks) would be recalled. For the E20 and NE20 conditions, the rank-order correlations between the IAS index and the number of times each category was represented in recall were .01 and .24, respectively. Neither of these correlations is significantly different from zero. Thus, despite the ability of the IAS index to discriminate between the E and NE categories and to predict within-category recall, the index provided no clue to account for the divergence of the chunk-recall means.

EXPERIMENT II

Aside from the nonsignificant differences between the U10, NE10, and E10 chunk-recall means (where a ceiling effect exists), the data of Exp. I did not support the experimental hypothesis. In a further attempt to account for the divergence, attention was focused on the time variable. While list length, measured in words, varied from 10 to 70 over the nine conditions, the presentation rate remained constant at 3 sec/word. Thus, 60 sec. were required to present the 20 words of the U20 list while 210 sec. were required to present the 70 words of the E20 and NE20 conditions. It seemed reasonable to assume that the longer the total presentation time the greater the influence of factors, such as implicit rehearsal, which would tend to increase both word and category recall. Experiment II was undertaken to assess the relation between chunk recall and total presentation time.

Method

Three groups of *S*s were tested. The *S*s were male undergraduate students attending the summer term at the Pennsylvania State University. One group ($N = 29$) received the U20 list at a rate of 10.5 sec/word. Thus, the total time to present the 20 words (210 sec.) was equivalent to the time required when the 70 words of the E20 and NE20 lists were presented at a 3-sec. rate. To distinguish this group from the previous U20 group, it was labeled U20(210). The second and third groups ($N = 19$ and 16, respectively) received the E20 and NE20 lists at a rate of 1 sec/word. In this manner, the total time to present the 70 words of these lists (70 sec.) was roughly equivalent to the 60 sec. required when the words of the U20 list were presented at a 3-sec. rate. These groups were labeled E20(70) and NE20(70) to differentiate them from the previous groups.

Each *S* was provided with one of the individual, manually-operated memory drums which were mentioned earlier.[3] The *S*s exposed the words, one after another, in time with a tape recorded signal. Because of the manual nature of this technique, a presentation rate of 1 sec/word was used for the E20(70) and NE20(70) conditions rather than the .86 sec/word which would be required to present the 70 words in 60 sec. It was felt that a manual presentation rate of less than 1 sec. would be too difficult a task for *S*s. The 10-sec. time differential was not considered to be crucial for the evaluation of the results. Even so, a number of *S*s failed to pace themselves correctly and either finished too soon or failed to expose all the words of the list. Only the data of those *S*s who successfully accomplished the word presentation will be presented here. In all other respects, the data collection was the same as for Exp. I.

Results

The slow presentation rate for the U20(210) condition resulted in a mean word recall of 13.00 ($SD = 2.91$). This mean was compared with the chunk-recall means of the E20 and NE20 conditions. The analysis of variance indicated no significant difference ($F < 1.0$). The fast presentation rate for the E20(70) and NE20(70) conditions resulted in a decrease in both word and chunk recall. The mean word recalls for the E20(70) and NE20(70) conditions were 22.74 ($SD = 8.84$) and 15.38 ($SD = 4.87$), respectively. The chunk-recall means for the E20(70) and NE20(70) conditions were 9.26 and 8.69 with *SD*s of 2.12 and 2.14, respectively. When these chunk-recall means were compared with the mean word recall of the U20 condition, the analysis of variance again indicated no significant difference ($F < 1.0$). The means for the U20(210), NE20(70), and E20(70) conditions are shown as separate points in Fig. 1. Based on these results, it may be concluded that the divergence of the chunk-recall means shown in Fig. 1 is a function of the differences in total presentation time.

---

[3] On the basis of the earlier study (Cohen, 1963), no significant differences were found between group testing with the manually-operated memory drums and individual testing with a standard memory drum.

Further evidence for the importance of total presentation time was obtained by the comparison of the number of three- and four-word categories represented in recall. For the E20(70) and NE20(70) conditions, the total presentation time for the three-word categories was 30 sec. compared to 40 sec. for the four-word categories. Similarly, for the E20 and NE20 conditions, the total presentation time was 90 sec. and 120 sec. for the three- and four-word categories, respectively. If total presentation time is a sensitive variable it would be expected that more four-word than three-word categories would be represented in recall. Table 3 presents the mean number of three- and four-word categories represented in recall for each of the category conditions of Exp. I and Exp. II. In every case more four-word categories were recalled. Because of the discrete nature of the data and the restricted range of scores, the differences between the recall of three- and four-word categories were evaluated by means of the nonparametric sign test. The resulting p values, as shown in Table 3, leave little doubt of the overall superiority of recall of four-word chunks.

TABLE 3

Mean Number of Three- and Four-Word Chunks Recalled

Cond.	Chunk Size		p
	Three-Word	Four-Word	
NE10	3.27	4.13	.046
NE15	3.73	6.20	.004
NE20	5.13	7.20	<.001
NE20(70)	3.63	5.06	.151
E10	4.13	4.53	.113
E15	5.13	6.00	.059
E20	5.93	6.87	.046
E20(70)	4.16	5.11	.059

Note.—p values are probabilities based on the sign test. In assessing differences for the E15 and NE15 conditions, the number of three- and four-word categories recalled per *S* were converted into percentages since there were eight four-word and seven three-word categories.

Discussion

When equated for total presentation time, the number of chunks recalled from the 20-category lists did not differ significantly from the number of words recalled from the list of 20 unrelated words. The demonstration that more four-word than three-word categories were represented in recall provided additional evidence for the importance and sensitivity of total presentation time as an independent variable. In view of these findings, if the original hypothesis is amended to take into consideration total presentation time (or equivalently, time per chunk), the investigation provides strong support for Miller's (1956) general statements concerning verbal recoding and the processing of information in immediate memory.

One aspect of the study deserves special comment. At a presentation rate of 3 sec/word, each three-word category required 9 sec. to present while each four-word category required 12 sec. Recall, however, that the words of the categorized lists were presented in a random fashion. Thus, the time required to present the words of a given category did not occur successively but occurred as three or four discrete 3-sec. exposures separated in time by varying numbers of words of other categories. It might be expected, therefore, that the longer the list length (in words) the greater the role of factors such as interference and differential forgetting which would, in general, tend to reduce the average word recall per category. No such tendency was observed. For the E10, E15, and E20 conditions, the average recalls of words of three-word categories were 2.63, 2.55, and 2.47, respectively. For the four-word categories, the average recalls were 3.40, 3.38, and 3.47. Similarly, for the three-word categories of the NE10, NE15, and NE20 conditions, the average word recalls were 1.92, 2.00, and 1.77 while for the four-word categories the values were 2.52, 2.60, and 2.44. If these means are converted into percentages of total possible category recall, a striking consistency is revealed. Despite the variations in list length and category size, each recalled E category was represented by approximately 85% of its words. Each recalled NE category was represented by approximately 63% of its words. This overall stability of within-category word recall for both the E and NE categories is consistent with the assumption that the categorized word lists are handled in immediate memory as so many chunks of information rather than as so many individual words.

Finally, mention should be made of the increase in word recall for the U20(210) compared to the U20 condition. Murdock (1960) amassed considerable data which showed that recall, after a single presentation of a list, is a linear function of total presentation time. As indicated previously, Murdock's formula predicted the mean recall for the U conditions with some accuracy—when the presentation rate was 3 sec/word. For the U20(210) condition, however, the predicted mean recall of 18.70 was considerably higher than the actual mean recall of 13.00. The slowest presentation rate used by Murdock in developing his formula was 4 sec/word whereas the U20(210) condition of this study utilized a rate of 10.5 sec/word. This suggests the possibility that recall of lists of unrelated words may not be a linear function of total presentation time when fairly slow presentation rates are used.

REFERENCES

Cohen, B. H. An investigation of recoding in free recall. *J. exp. Psychol.*, 1963, **65**, 368–376.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Murdock, B. B., Jr. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received September 8, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 235–240

EFFECT OF INSTRUCTIONS AND UCS STRENGTH ON EXTINCTION OF THE CONDITIONED GSR[1]

DELOS D. WICKENS, CHARLES K. ALLEN, AND FRANCES A. HILL

Ohio State University

4 experimental groups received 10 tone-shock pairings at a 500-msec. interstimulus interval, and 2 control groups received corresponding unpaired tones and shocks. 2 experimental groups received strong and 2 others weak shocks with the tone, the converse shock being interspersed between pairings. Before extinction was begun, half of these groups were told that no further shocks would be given and their corresponding group was not so informed. The data during extinction indicate that: (a) Both controls were inferior to their corresponding experimentals indicating the presence of an association factor. (b) The 2 controls differ in response magnitude suggesting that 1 effect of instructions is to lower drive. (c) Instructions do not immediately eliminate responses to the CS in the experimentals, but rather result in a faster rate of extinction. (d) No clear evidence for an effect of UCS strength upon conditioning was found.

This study was undertaken to investigate the effect on the extinction of the GSR of informing *S*s that no further shocks would be given. This type of investigation is not new. In 1937 Cook and Harris demonstrated that the GSR is markedly reduced if *S* is told that shock will no longer follow the CS, and almost concurrently Mowrer (1938) reported a similar finding. More recently, Silverman (1960) has added a complication to the earlier conclusion by showing that the interstimulus interval interacts with the effects of instructions. Certainly the conclusions drawn in the two early studies tend to leave the impression that the response to the CS is controlled almost totally by verbal processes, but this is less true of the Silverman study and does not seem to be the case for other types of conditioned responses (Hartman & Grant, 1962; Lindley & Moyer, 1961).

The present investigation is an effort to re-examine the effect in GSR conditioning in somewhat greater detail by using certain controls which were lacking in the earlier studies and by evaluating the process across extinction trials. It included groups conditioned to a weak or a strong UCS so that the additional hypothesis (Spence, 1960) that strength of conditioning to an aversive stimulus is a function of UCS strength could also be tested.

METHOD

Subjects.—A total of 141 elementary psychology students participated in the experiment, 21 of whom were discarded because of base levels above 90,000 ohms, failure to react to the UCS, or experimental errors. The remaining 120 *S*s were equally distributed in the six groups of the experiment. Their assignment to groups was made randomly by drawing from a stack of cards on which the six group designations appeared.

Apparatus.—The time relations and duration of the CS and UCS were controlled by Tektronix equipment and calibrated by a Hewlitt-Packard electronic counter. The CS was generated by a Jackson oscillator and the shock UCS by a dc power pack. The GSR was recorded from a Fels dermohmeter using the standard Fels zinc electrodes and a jelly composed of bacto agar, zinc sulphate, and distilled water. The *S* and the control equipment were in adjacent rooms and *S* wore

---

[1] This research was supported by National Science Foundation Grant G 19606 to the senior author.

Willson sound barrier ear muffs containing earphones so as to further reduce extraneous noises. He sat in a dental chair with the room dimly illuminated by a General Electric night light.

Procedure.—After *S* entered the experimental room he stood by a table while the GSR electrodes were placed on the palm and back of the left hand and the shock electrodes on the back and middle finger of the right hand. He then sat in the experimental chair and *E* attached leads to each electrode. The purpose of first placing the electrodes and then attaching leads was to convey the accurate notion to *S* that shock could occur only if the leads were affixed to the electrodes. The *S* was then told that he would hear a tone in the earphones and receive a shock but there was nothing he needed to do. The *E* then left the room and presented three shocks alone, the door between the two rooms being open during this time. The values of these three successive stimulations were 1.5, 1.6, and 1.7 ma. Then the door to *S*'s room was closed and one of the three different training procedures was begun.

The six groups of the experiment are distinguished by the method of training and whether or not certain instructions were given prior to extinction. The three different training procedures were as follows.

Training.—Conditioning: Strong Shock. The CS was associated with a strong shock for 10 trials and 10 weak shocks were interpolated between trials. The interstimulus interval was 500 msec., the tone remaining on and terminating with the 100-msec. UCS. No interpolated shock preceded or followed a CS by less than 20 sec., and no more than three reinforcements occurred successively.

Conditioning: Weak Shock. This procedure was identical to that described in the previous paragraph save that the weak shock was associated with the CS and the strong shock was interpolated.

Control. For this group a total of 10 tones, 10 strong and 10 weak shocks were presented in the same total time required for the training of the above two groups. No tone preceded or followed a shock by less than 15 sec.

Extinction.—Prior to the beginning of the extinction series, *E* entered *S*'s room and made one of the two following statements:

> There will be no more shocks, so I will remove these electrode leads. [The leads were then pulled out of the electrode sockets and hung on the wall.]
>
> I just wanted to make sure that the GSR electrodes are still secure. [The electrodes were not touched but merely examined visually.]

Half of the 40 *S*s in each training group received the first of these statements, and half received the second. Those given the no-shock instructions will hereafter be referred to as the *Instructed groups.* The *E* then left the room and 30 sec. after the door was closed an extinction series of five trials was begun.

The milliampere level of the weak shock was 1.0 and that of the strong was 2.5. However, in order to adapt *S* to the strong shock, it was raised by .2-ma. steps on successive stimulations from 1.7 until the 2.5 level was reached. The intertrial interval during training ranged randomly from 15 to 50 sec. The mean intertrial interval was 33 sec. The five extinction trials were spaced 30 sec. apart.

The GSR was recorded visually from a dial on the face of the instrument and was defined as the response which occurred within approximately 2 sec. of stimulus presentation. The scores were transformed according to the formula:

$$\log\left[\left(\frac{1}{R_a} - \frac{1}{R_b} \times 10^8\right) + 1\right]$$

where R_b is the resistance at the moment of initiation of the response and R_a is the resistance at the peak of the response. The 1,000-cps CS was 60 db. re .0002 dyne/cm^2.

Following the experiment *S*s were asked if they believed *E*'s statement that no more shocks were to be given. Slightly less than 1 in 5 *S*s reported that they were not confident that the shock would not be presented.

RESULTS

Base level.—The mean base level, the R_b term in the transformation equation, for all the groups on the last training and all extinction trials is presented in Fig. 1. It is apparent from an examination of this figure that the level dropped following *E*'s entry into *S*'s room. A Trials × Groups analysis of variance of the last training and first extinction trial indicated that the drop is significant, $F\,(1, 114) = 32.46, p < .01$, error (w) = 1090.09, but that the groups did not differ significantly from each other on either trial nor was there a significant interaction between groups and trials.

A Trials × Groups analysis was also conducted across extinction trials and again no group difference was found, but the effects of Trials were significant, $F\ (4, 456) = 117.44, p < .01$, error (w) = 794.48, as also was the interaction between Groups and Trials, $F\ (20, 456) = 4.35, p < .01$, error (w) = 794.48. It is apparent that this interaction is primarily due to the smaller amount of rise in resistance of the two noninstructed conditioning groups.

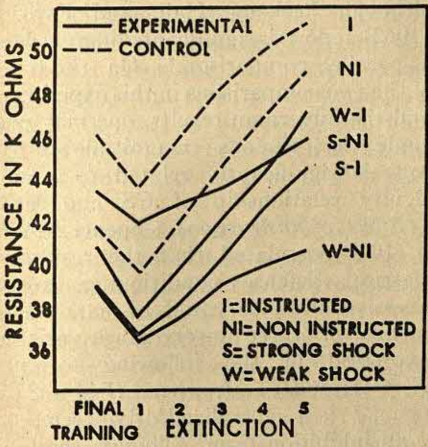


FIG. 1. Mean base level in thousand ohms for each group on the last acquisition trial and each extinction trial.

Response during extinction.—The mean response magnitude for each group on each extinction trial is presented in Fig. 2. A Trials × Groups analysis of variance for these data was also performed. For the purpose of this analysis, the dimension of UCS strength was assumed to have three values, strong, weak, and zero, the zero value being assigned to the control groups since no shock was presented in close temporal contiguity to the tone. All sources of variance were significant at $p < .025$ save the triple order term and the interaction between Instructions and UCS Strength. Trials, as a main effect, was not tested because the obvious differences shown in Fig. 2 between the controls and experimentals make such a test meaningless. The significant UCS effect, as will be demonstrated later, was attributable solely to the control groups.

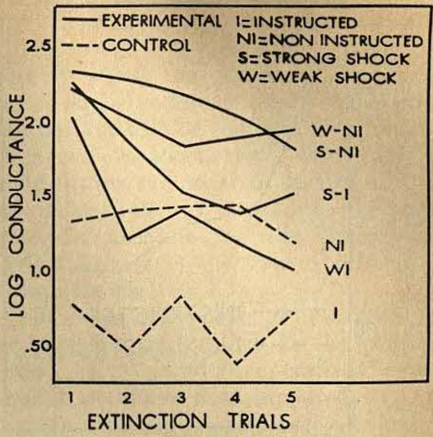


FIG. 2. Mean response magnitude to the CS for all groups on each extinction trial.

A second Groups × Trials analysis was conducted using only the four experimental groups and this is presented in Table 1. The main effects of Trials and Instructions continue to be significant and an interaction of

TABLE 1

TREATMENT × TRIALS ANALYSIS OF VARIANCE FOR THE EXPERIMENTAL GROUPS

Source	*df*	*MS*	*F*
Between *S*	79		
UCS strength	1	6.5917	2.67
Instructions (I)	1	25.3504	10.26**
UCS × I	1	1.0220	
Error (b)	76	2.4712	
Within *S*	320		
Trials (T)	4	5.1150	21.28**
T × UCS	4	.2456	1.02
T × I	4	1.1115	4.62**
T × I × UCS	4	.4990	2.08
Error (w)	304	.2404	
Total	399		

** $p < .01$.

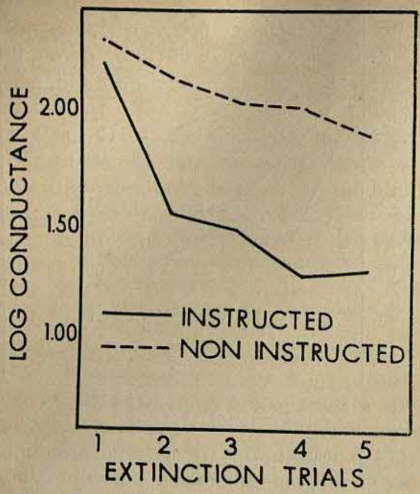


FIG. 3. Extinction curves for the Instructed and Noninstructed groups with Weak and Strong shock groups collapsed.

Trials and Instructions was found. Since there was no evidence that UCS strength was a significant source of variance, this dimension was collapsed and *t*'s were run between the Instructed and Noninstructed groups for certain trials. No significant difference was found between the two groups on the first trial, but they differed at $p < .02$ on the last trial. Furthermore, performance on the last trial for each group differed significantly ($p < .01$) from performance on the first trial. The extinction curves for these combined groups are presented in Fig. 3.

Finally, another Trials × Groups analysis was conducted on the two control groups and the only significant result obtained was that for Groups, $F(1, 38) = 12.75$, $p < .01$, error (b) $= 2.0537$. In other words, neither of these groups showed a significant decrement in performance during the five extinction trials.

DISCUSSION

Effect of temporal contiguity.—The performance of the control groups was primarily responsible for the significant effect of UCS strength in the Trials × Groups analysis of all six groups, for no such effect was found in the analysis of the Experimentals alone. Additional analyses were performed in which each *S*'s scores were summed across the five extinction trials and the means of these scores obtained for each group. The Control group mean for the Instructed group was then compared by *t* tests to the two appropriate Experimental groups and the same operations were performed on the data for the Noninstructed groups. In all instances a $p < .02$ was obtained. The results are consistent with those obtained by Silverman for similar groups, although Prokasy, Hall, and Fawcett (1962) report insignificant differences in a parallel comparison.[2] On the basis of the two comparisons in this experiment and the Silverman results, one may conclude that response magnitude to the tone is significantly related to a contiguity relationship of CS and UCS.

Effect of instructions.—Spence (1960, p. 109) postulated that the results of classical defense conditioning to an aversive stimulus could be handled by means of three intervening variables expressed in the following formula: $E = (H \times D) - I$, where H is a habit term, D is a generalized drive term, and I an inhibition term. Though the GSR cannot readily be placed in the category of defense conditioning, the effect produced by contiguity will be considered as analogous to the H in the classical defense formulation.

Control group.—The control group results are readily interpreted within this formula by assuming that the tone has a small capacity for eliciting the GSR. Thus, the value of H, or perhaps more accurately some unconditioned tendency to evoke the GSR (Sokolov, 1960), is slightly but equally positive for both groups. The D term, as Spence (1960) uses it, is assumed to be "based upon fear responses of different strength that become conditioned to the cues of the experimental situation [p. 162]." The occurrence of shock in the experimental context should, therefore, give rise to a D and this multiplies with the H term to determine the positive value of E. One needs only to assume that the verbal statement that no further shock will be given produces a lower D level for the instructed groups to account for the difference between the two groups.

---

[2] In a private communication Prokasy has reported that their failure to obtain differences between their groups was probably a function of the intertrial intervals used. In a subsequent experiment they found a difference between controls and experimentals with intertrial intervals comparable to those used in the present experiment, but not under massed conditions.

Neither group shows a significant decline in response during the extinction trials. For the instructed group this fact is not surprising for it may be close to the zero value from the first trial onward. Obviously, however, the noninstructed group could decline to the level of the instructed group but it does not do so.

Experimental groups.[3]—As with the controls, the overall level of response is greater in the noninstructed group than in the instructed, but they differ from the controls in several ways: (*a*) a difference in overall height of the curves, (*b*) the fact that the experimentals show an extinction effect, and (*c*) no significant difference between the two experimentals on the first extinction trial.

The difference in overall height may be attributed to the development of H during training for the experimentals and the failure for this to occur in the controls. Assuming for the time being equality of D for experimentals and controls, the value of the $(H \times D)$ would be larger for the experimentals.

The I term will obviously account for the decline in the experimental curves. Since there is no evidence for response decrement in the control groups across nonreinforced trials, it is necessary to assume that increments in I did not occur for these groups. Spence (1960, p. 97) assumes for instrumental reward conditioning that inhibition does not develop unless the S-R event had been previously reinforced and the same assumption will be made in the present instance. In short, the I term cannot build up for the controls because of the low value of H.[4] This analysis does not however, account for the differential extinction rate of the two experimental groups themselves.

Since the drive concept was used to account for instructions in the control groups, and since a generalized drive should also have been acquired by the experimentals, this same scheme should be applied to these groups. Assuming that the sole difference between these groups is in the value of D, one would obtain two extinction curves which differ in height, as indeed the two curves do, but the two curves would be parallel to each other and the interaction between trials and groups would not have been obtained. Hence, although one may tentatively assume that the two groups differ from each other with respect to D in the same manner as the controls, an additional variable is required to account for the divergent extinction curves so clearly illustrated in Fig. 3. This can readily be handled by the I term if the assumption is made that I is not only a function of nonreinforced trials, but also that the *increment* in I following a nonreinforced trial is a function of an attitude engendered in S by the instructions. This could be expressed by a term such as $I = E_0(1 - e^{-aT})$ where T represents nonreinforced trials, a is a constant which varies with S's attitude, and E_0 is a constant equal to the value of $(H \times D)$ at the time the extinction trials begin. Differences in the value of these constants would result in marked differences in value of I once T is brought into operation. The aT term implies that experiences subsequent to the verbalizations are important determiners of the influences of instructions on behavior, and that verbal instructions alone have no effect upon this variable. The E_0 term expresses the above mentioned view that the course of extinction is a function of previous reinforcements.

---

[3] Since no significant effect of UCS strength was found, this discussion is concerned with the Instructed and Noninstructed groups collapsed with respect to UCS strength.

[4] Although it is appropriate to omit the I term to account for the course of behavior in the experiment itself, its omission would imply that the Noninstructed controls would never decline to the level of the Instructed controls. It seems unlikely that such would be the case over an extended period of time. This conjecture suggests that there may be an inhibition term which operates directly on D.

Judging by the failure to find differences between the instructed and noninstructed experimentals on the first trial, the D term does not seem to be as important for the experimentals as for the controls. It is possible that D is of lower value for the experimentals than for the controls because for the controls shock is never associated with a specific stimulus and becomes exclusively a context produced effect, while for the experimentals the emotional or anxiety reaction has an opportunity to become tied to a specific phasic stimulus and is evoked to a lesser degree by the context.

Effect of UCS intensity.—The results insofar as variations in UCS intensity for the two experimentals are somewhat ambiguous. In neither base level measure nor response magnitude measure does UCS intensity or its interaction with other variables reach the usually acceptable .05 level. An analysis of covariance against base level was conducted on the data in an effort to give UCS strength a more favorable chance to demonstrate its effect, but the results of this analysis left the F value for UCS strength essentially unchanged.

In order to determine whether or not the two levels of the UCS were differentially effective, a comparison was made of the response to the shock on the last shock alone trials for the experimental groups. The mean response to the weak shock was 2.0583 units, and the mean response to the strong was 2.6189. The $t(63)$ value for this difference is 4.28, which yields $p < .01$.

In the field of eyelid conditioning, Spence, Haggard, and Ross (1958), using the same technique for equating drive as that employed in the present experiment, found evidence for superior conditioning in the strong UCS group and this finding is used to support the view that H is a function of UCS strength in defense conditioning. While the means in the present experiment are ranked in the order strong to weak for both pairs of the experimental group, the p level for the intensity effect obtains only the .16 level, and the results permit no definitive statement concerning the role of UCS strength in GSR conditioning.

REFERENCES

COOK, S. W., & HARRIS, R. E. The verbal conditioning of the GSR. *J. exp. Psychol.*, 1937, 21, 202–210.

HARTMAN, T. F., & GRANT, D. A. Effects of pattern of reinforcement and verbal information on acquisition, extinction, and spontaneous recovery of the eyelid CR. *J. exp. Psychol.*, 1962, **63**, 217–226.

LINDLEY, R. H., & MOYER, K. E. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *J. exp. Psychol.*, 1961, 61, 82–88.

MOWRER, O. H. Preparatory set (expectancy): A determinant in motivation and learning. *Psychol. Rev.*, 1938, **45**, 62–91.

PROKASY, W. F., HALL, J. F., & FAWCETT, J. T. Adaptation, sensitization, forward and backward conditioning and pseudo-conditioning of the GSR. *Psychol. Rep.* 1962, **10**, 103–106.

SILVERMAN, R. E. Eliminating a conditioned GSR by reduction of experimental anxiety. *J. exp. Psychol.*, 1960, **59**, 122–125.

SOKOLOV, E. N. Neuronal models and the orienting reflex. In Mary A. B. Brazier (Ed.), *The central nervous system and behavior.* Madison, N. J.: Madison Printing Co., 1960.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411.

(Received September 10, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 241–244

EFFECT OF PATTERN OF REINFORCEMENT ON THE CONDITIONED EYELID RESPONSE[1].

GEORGE E. PASSEY AND DAVID L. WOOD

University of Alabama

4 groups of 10 Ss each acquired conditioned eyelid responses to a 1024-cps tonal stimulus delivered at 40 db. above threshold. Each of the 4 groups was given the same mean intensity of reinforcing air puff over a 50-trial acquisition series. Pattern of reinforcement differed for each of the groups with the employment of ascending, constant, variable, and descending UCS magnitudes. The groups showed no differences in number of anticipatory responses during the acquisition series though the slopes of the acquisition curves were significantly different from one another. In order of slope the groups were ascending, constant, variable, and descending, with the ascending group showing the steepest acquisition curve. No sex differences in rate of acquisition were found.

Magnitude of the reinforcing air puff has been shown to be an effective determiner of performance in eyelid conditioning (Passey, 1948; Spence & Taylor, 1951). More recently Prokasy, Grant, and Meyers (1958) found that intensity of the UCS was a significant factor in the acquisition of the eyelid response, and that UCS intensity exerted its influence throughout 2 days of acquisition training, thus influencing both rate and final level of conditioning.

Passey and Burns (1962) have shown that randomly variable magnitude in pressure of the reinforcing air puff for eyelid conditioning under 100% reinforcement, leads to lower performance as the variability of the reinforcing air puff increases. Three degrees of variability were employed and the effect was demonstrated for acquisition trials only. The present study was carried out to assess the effect of progressive incremental and progressive decremental reinforcement patterns upon acquisition of the conditioned eyelid response. A control group conditioned under constant reinforcement at the mean of the incremental and decremental series was employed. A randomly variable reinforcement group was included to provide comparison with the finding of Passey and Burns.

METHOD

The conditioned eyelid response was developed in four groups of *S*s. Each *S* received 50 paired CS-UCS trials in which a 1,024-cps tone, 40 db. above threshold, served as the CS. The UCS was an oxygen puff to the cornea and UCS series were composed of five delivery pressures of 3.2, 4.4, 5.5, 6.8, and 8.0 psi at the eye. The diameter of the delivery aperture was $\frac{5}{32}$ in. and was located 2 in. horizontally from *S*'s eye. Each of the four groups was conditioned at an average gauge pressure of 5.5 psi.

The ascending series group received an initial pressure of 3.2 psi during the first 10 trials, 4.4 psi for the second block of 10 trials, and so on, terminating with the fifth block of 10 trials at 8.0 psi. The constant group received the UCS at 5.5 psi for all trials. The UCS for the variable series group consisted in application of all five pressures varied in a random manner and generating an overall mean of 5.5 psi. The descending series group was the inverse of the ascending series group with the first 10 trials involving a UCS intensity of 8.0 psi. The standard deviation for the ascending, random, and descending

---

[1] This research was supported in full by a grant from the Research Committee of the University of Alabama.

groups was 1.9 psi while that for the constant group was .9 psi.

The duration of the CS was 700 msec. and the duration of the UCS was 50 msec. The separation between the onset of the CS and the onset of the UCS was 680 msec. The UCS overlapped the last 20 msec. of the CS and extended beyond its terminus for 30 msec. The intertrial interval was variable in a sequence which repeated each 11 trials. The median intertrial interval was 19 sec. and the range was from 12 to 26 sec.

Apparatus.—Eyeblinks and the onset and termination of the CS and UCS were recorded on a Brush oscillograph. Responses were detected by use of a Hunter motions amplifier. The CS was produced by a Maico audiometer and was delivered via headphones. The air puff was provided by an oxygen cylinder with reducing valves for pressure adjustment. Five electrically operated valves delivered air puffs. Selection of one of the five valves would determine UCS magnitude on any trial. A cam timer controlled CS and UCS presentations and duration. A second cam timer effected the variable intertrial intervals. An air conditioner maintained a constant temperature in the experimental chamber and also served to mask extraneous auditory stimuli.

Procedure.—The steps in stimulation were as follows:

1. Each *S* was instructed that the investigation concerned eyelid conditioning. He was asked to relax as much as possible and to make no attempt to control the movement of his eyelids. He was told that the experiment would last for about 30 min. and was asked to look at an illuminated milk glass screen directly ahead at a distance of 6 ft. for the duration of the session.

2. An auditory threshold was then established by the methods of limits and the CS intensity control was set.

3. The response recorder was then attached to the eyelid by means of a strip of adhesive tape. The air conditioner was started and *E* left the room.

4. The *S* was then stimulated: twice with the CS alone; three times with the UCS alone; and twice again with the CS alone.

5. Following the preconditioning series, 50 paired CS-UCS presentations were administered.

6. Upon termination of the stimulation series, *S* was interrogated so as to discover whether any extraneous cues had influenced his behavior.

Subjects.—Four groups of 10 *S*s each were employed. The *S*s were recruited from classes in introductory psychology as partial fulfillment of course requirements. Of the 52 *S*s recruited, 12 were rejected. There was no systematic attachment of rejected *S*s to any treatment group. The criterion for rejection of an *S* as a sensitized responder was that he respond to either of the final unpaired CS presentations. Any response occurring during the initial 200 msec. of the CS presentation was excluded from the data. An *S* who had 50% or more of his anticipatory responses with a latency less than 200 msec. was eliminated from consideration. Of the 12 *S*s rejected, 5 were lost due to procedural errors and the remaining 7 eliminated on the sensitization or latency criterion. Responses were not scored as CRs unless the response amplitude equaled or exceeded 2 mm. The sex composition of each group was six females and four males.

RESULTS AND DISCUSSION

Conditioning records of the *S*s were analyzed by counting the number of anticipatory responses during the 50-trial acquisition series. These data are summarized for the various experimental treatments in Table 1 where they are grouped for blocks of 10 trials.

The data were analyzed following the technique suggested by Edwards (1960). The results of this analysis are presented in Table 2. Test for differential effects of the experimental treatments yielded an $F(3, 36) < 1.0$, $p > .05$. We conclude that the amount of conditioning in each of the experimental groups is essentially the

TABLE 1

AVERAGE NUMBER OF ANTICIPATORY RESPONSES FOR BLOCKS OF 10 ACQUISITION TRIALS

Group	Blocks of 10 Trials				
	1	2	3	4	5
Ascending series	2.4	5.7	6.4	7.3	7.6
Constant series	3.9	6.1	6.9	7.1	7.4
Variable series	4.3	4.5	6.2	6.4	6.1
Descending series	4.5	4.9	4.9	5.4	5.4

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF THE NUMBER OF ANTICIPATORY RESPONSES DURING 50 PAIRED CS-UCS PRESENTATIONS

Source	*df*	*MS*	*F*
Experimental treatment (ET)	3	14.46	<1
Error (a)	36	53.83	
Trials (T)	4	55.99	21.70***
ET × T	12	6.41	2.48**
Error (b)	144	2.58	

** $p < .01$.
*** $p < .001$.

TABLE 3

SUMMARY OF TREND ANALYSIS FOR ACQUISITION DATA

Source	*df*	*F*
Linear component	1	74.89***
Curvilinear component	1	11.70***
Linear trend	3	6.50***
Curvilinear trend	3	1.58
Error (*MS*)	144	(2.58)

*** $p < .001$.

same. Since the number of trials did not allow the response to reach asymptote we do not know what might have occurred had the acquisition series been extended.

Utilizing Edwards' suggestion the variance was separated into linear and curvilinear components. The results of this analysis are presented in Table 3. For linear components F (1, 144) = 74.89, $p < .001$. An examination for the presence of curvilinear components yielded F (1, 144) = 11.7, $p < .001$. Since significant linear and curvilinear components were found, slope differences and curvature differences were examined. The F (3, 144) for linear trend differences was 6.5, $p < .001$. We conclude that the slopes of the performance curves are not parallel and that there was a differential rate of acquisition among the experimental groups. The test for differences in curvature yielded F (3, 144) = 1.58, $p > .05$. The hypothesis of no difference in curve form is accepted.

Evaluation of the differences among individual trend pairs was carried out using the technique suggested by McNemar (1962). The trends not only showed an overall divergence as indicated in the previous analysis but additionally all of the slopes differed from each other. All comparisons yielded probabilities of chance occurrence <.05. Each experimental treatment reflected its own unique acquisition rate.

The relatively steep slope for the ascending as compared with the randomly variable and decremental magnitude groups, despite their equal variability, suggests the importance of patterning of reinforcement as a determiner of the slope of the acquisition curves. This finding is related to that of Flugel (1928) who found that incremental reward magnitude delayed the appearance of negative acceleration in his performance curves until long after it would normally appear with magnitude of reward held constant. The inferior rate of conditioning for the randomly variable as compared with the constant reinforcement magnitude group agrees with the finding of Passey and Burns (1962).

A comparison of performance as a function of sex yielded $t(38) = 1.71$, $p > .05$. The possibility of differential conditioning between the sexes remains an unsettled issue. McAllister (1953) and Reynolds (1958) have found that females condition more readily than males.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

FLUGEL, J. C. Practice, fatigue and oscillation: A study of work at high pressure. *Brit. J. Psychol. monogr. Suppl.*, 1928, No. 13.

McALLISTER, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–422.

McNEMAR, Q. *Psychological statistics.* New York: Wiley, 1962.

PASSEY, G. E. The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.

PASSEY, G. E., & BURNS, T. C. The influence of variable reinforcement upon acquisition of the conditioned eyelid response. *Psychol. Rep.*, 1962, **11**, 547–552.

PROKASY, W. F., GRANT, D. A., & MEYERS, N. A. Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55**, 242–246.

REYNOLDS, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55**, 335–341.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

(Received September 15, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 245–254

DETERMINANTS OF SELF-REINFORCEMENT IN HUMAN LEARNING[1]

FREDERICK H. KANFER AND ALBERT R. MARSTON

University of Oregon Medical School *University of Wisconsin*

In 3 separate experiments frequency and accuracy of self-reinforcing responses (SRs) were investigated as a function of (a) stage of prior learning, (b) instructions about stringency of criteria for SR, and (c) similarity of discriminative and reinforcing stimuli in acquisition and test for SR. It was found: (a) learning to a higher criterion resulted in more frequent and more accurate SR responses, (b) instructions establishing more lenient criteria for SR led to more frequent but more incorrect SRs, and (c) when the test list differed from the acquisition list, SRs were lower in frequency but higher in proportion of incorrect SRs. Presence or absence of the original reinforcer did not affect SR. The SR data were also related to occurrence of correct responses on the test.

In two previous studies *S*s learned a verbal discrimination to a low criterion. The *E* reinforced correct responses by turning on a light. The *S*s were then instructed to take over *E*'s control of the reinforcing stimulus by turning on the light themselves when they thought their responses were correct. This procedure was called self-reinforcement (sr). It was found that the frequency and accuracy of self-reinforcing responses (SRs) increased with more learning (Kanfer, Bradley, & Marston, 1962). In a second study using the same procedure, discrimination of the previously learned responses under sr was found to be significantly better than in extinction and lower than in continued *E*-controlled reinforcement (acquisition) (Marston & Kanfer, 1963). These findings suggest that SRs may affect performance on a task in the absence of external (*E*-delivered) reinforcement. The major purpose of the present study was to examine, in three separate experiments, the following parameters of the frequency and accuracy of SRs: Amount of learning on a discrimination task (Exp. I); instructions about stringency of criteria for SRs (Exp. II); similarity of the discrimination stimuli during acquisition and during the test under sr, and absence during sr of the reinforcing stimulus used in acquisition of the discrimination (Exp. III).

Frequency of correct responses on the discrimination task, usually the major dependent variable in most learning studies, was examined here only in relation to SRs. Accuracy of SRs was defined as administration of the reinforcing stimulus by *S* when his response on the discrimination task was correct. Incorrect SRs were defined as SRs following an incorrect discrimination; thus total frequency of SRs consisted of the number of all SRs, whether correct or incorrect.[2]

EXPERIMENT I

The purpose of this study was to examine the effect of amount of original discrimination learning on the number of subsequent SRs and on the number of correct responses (CRs) during sr following the original discrimination learning.

---

[1] This study was supported in part by Research Grant MH-06921-01, F. H. Kanfer, principal investigator, from the National Institute of Mental Health, United States Public Health Service. The authors wish to acknowledge the assistance of Marcia M. Bradley in all phases of this study.

[2] The sr procedure permits four possible combinations of CR and SR responses. A CR could be followed by an SR or nothing, and an incorrect discrimination response (non-CR) could be followed by an SR or nothing. In the present experiments an SR was classified as correct or incorrect only if it was actually made by *S*; failure to follow a CR with an SR, and absence of an SR following a non-CR were not included in the analyses.

Method

Subjects.—Thirty male college students served as *S*s.

Apparatus.—The *S* sat before a table separated from *E* by a 2 × 2 ft. board. At eye level in the board was a 6-in. square window for stimulus presentation. Beneath the window was a green signal light controlled by a timer set for 1 sec. and activated by a microswitch. The same microswitch also activated a signal light on *E*'s side of the board. The *S*'s task was to select the correct one among four nonsense syllables typed on 4 × 6 in. white cards. There were 10 cards, each with a different set of four nonsense syllables of medium association value (47% or 53%). The position of the correct stimulus was varied randomly from card to card, with each of the four positions used approximately equally often.

Procedure.—There were two parts to the task: Discrimination learning (Phase I), and self-reinforcement trials (Phase II). In Phase I, *S* was instructed to select one syllable on each trial and spell it. Ten trials constituted one block. On each block the cards were presented in different random order. The signal light was turned on whenever *S* selected the correct syllable. Card presentation continued until *S* reached the criterion set for his group.

In Phase II (self-reinforcement) *S* was given the following instructions before receiving 10 additional blocks of trials:

> I want to see how well you can do when you take over my job. I will no longer operate the experiment. This is the switch to turn on the green light. Try it once. Flash the light each time you have chosen the correct syllable. Now, when a card appears, quickly respond with the syllable which you think is correct. It is important that you do not spend any time in making your decision. After you have spelled the syllable, press the switch if you think you were correct. Do not press the switch unless you are pretty sure, that is unless you are fairly confident that you are correct.

The *E* never reinforced any response in Phase II.

Three groups of 10 *S*s each were given varying amounts of discrimination learning in Phase I. The *S*s in Group LL (low learning) continued acquisition to the end of the block in which a criterion of 5/10 correct was reached; Group ML (medium learning) learned to 7/10 correct; and Group HL (high learning) learned to 9/10 correct. The mean number of correct responses for Group LL on the last block of Phase I was 5.3; for Group ML, 7.4; and for Group HL, 9.0. Group LL took a mean of 5.5 blocks to reach criterion, Group ML took 10.9 blocks, and Group HL learned in 12.6 blocks.

Results

Figure 1 shows the response curves of the three groups for SRs and CRs. Frequency of SRs clearly increased with the level of discrimination learning and remained fairly stable over blocks in Phase II. The number of CRs also remained fairly constant in the absence of any reinforcement from *E*. The relative differences between mean number of SRs and CRs (Fig. 1) appear to decrease from Groups LL, ML, to HL, suggesting that more learning in Phase I resulted in closer correspondence of SRs and CRs. In all groups the SR curves remained higher than the CR curves, the latter approximating the criterion of the discrimination learning in Phase I.

The groups were compared by 3 × 10 repeated measures analyses of variance. As expected, CRs increase significantly with length of discrimination learning, F (2, 27) = 43.72. The analysis on SRs shows strikingly similar results. Frequency of SRs was significantly greater for the groups with more learning in Phase I, F (2, 27) = 6.11. The means per block for SR and CR, respectively, for Groups LL, ML, and HL were: 6.5 and 4.5; 8.1 and 7.1; 9.3 and 8.7. The mean proportion of incorrect SRs was .096 in HL, .154 in ML, and .374 in LL. The analysis on arc-sine

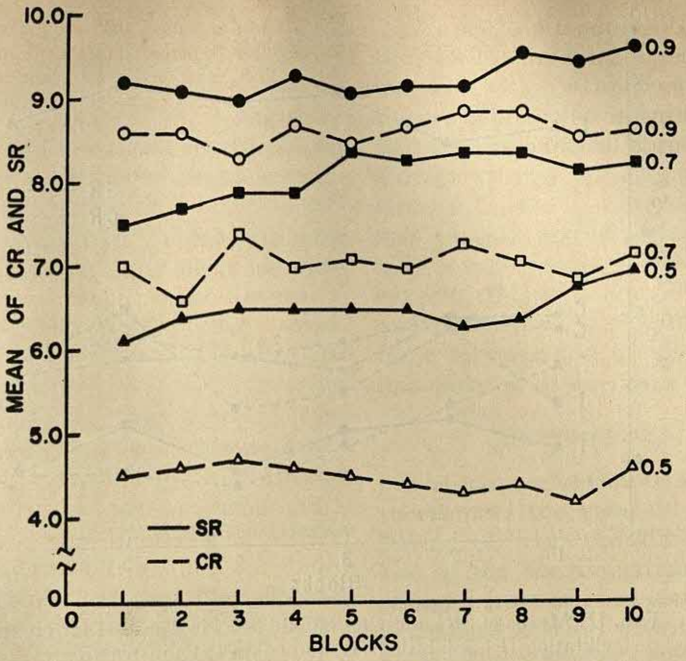


FIG. 1. Exp. I: Means of SRs and CRs on all blocks in Phase II for groups learning to a criterion of .5, .7, and .9 in acquisition.

transformations of the percentage of incorrect SRs indicates that *S*s with more learning in Phase I gave a significantly lower percentage of incorrect SRs, *F* (2, 27) = 11.76.

EXPERIMENT II

The purpose of this study was to investigate the effects of instructions designed to encourage or caution *S* about making SRs, on the same dependent variables as in Exp. I. In the present experimental procedure, *S*'s confidence about the accuracy of his responses is reflected directly by his SRs. Therefore it can be expected that the total frequency of SR would vary as a function of instructions which provide *S* with criteria for evaluating his response accuracy. If encouraged toward SRs he should maximize SRs, without regard to incorrect ones, while instructions cautioning *S* not to err should lower the total frequency of SRs but also minimize the frequency of incorrect SRs.

Method

Subjects.—Eighteen male college students were randomly assigned to three groups ($N = 6$ in each group).

Apparatus.—The same apparatus was used as in Exp. I.

Procedure.—The procedure was similar to that in Exp. I, with the following changes:

1. Three sets of instructions were used in an attempt to modify *S*'s criteria for evaluating the accuracy of his discrimination, as measured by SRs: (*a*) facilitating, (*b*) inhibiting, (*c*) control. At the beginning of Phase II, all *S*s received the same instructions as in Exp. I. Nothing more was said for the *control* (C) group. The two experimental groups received additional instructions. The *facilitation* (F) group was told that "the syllables are difficult to recall." They were encouraged to press the switch "even if you are not too sure whether you are actually right." It was stated that *S* can assume that

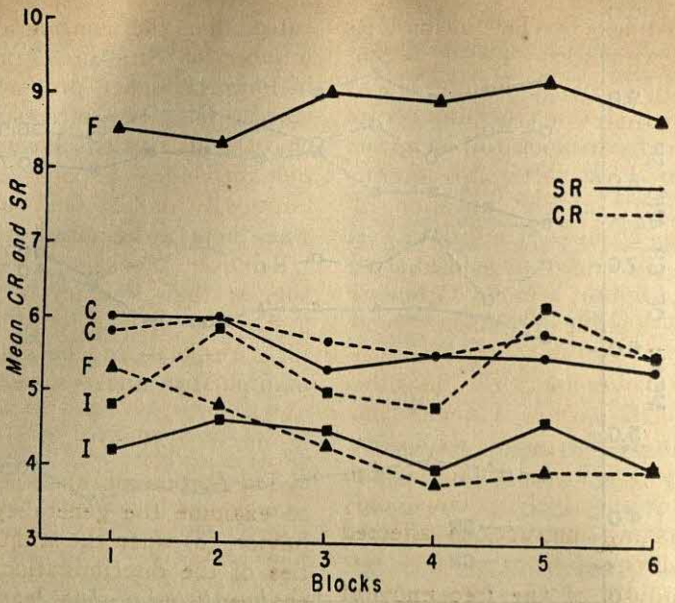


FIG. 2. Exp. II: Means of SRs and CRs on all blocks in Phase II for groups with facilitating (F), inhibiting (I), and control (C) instructions.

he is correct if there is only a little doubt and he is reasonably confident. The *inhibition* (I) group was told to be cautious and not to press the switch unless absolutely positive that they recalled the syllable. Further: "We are trying to see if students can carefully control their own rewards and not take them when they are not deserved."

2. In discrimination training (Phase I) a criterion of 6/10 was set for all groups. An analysis of variance revealed that the groups did not differ in trials to criterion or frequency CRs on the last block of discrimination learning.

3. Only six blocks of trials were given in Phase II, yielding 60 SR trials.

Results

The number of CRs and SRs in Phase II is shown in Fig. 2. Group F clearly showed a higher level of SRs than the other groups. The statistical analyses of these data are presented in Table 1. The variation in instructions did not affect the total number

TABLE 1

ANALYSES OF VARIANCE ON SRS, CRS, AND ARC-SINE TRANSFORMS OF PROPORTION OF INCORRECT/TOTAL SRS IN EXP. II

Source	*df*	CR		SR		ISR/TSR	
		MS	*F*	*MS*	*F*	*MS*	*F*
Groups (A)	2	17.13	2.59	185.95	11.83***	13.47	15.85***
*S*s in groups	15	6.62		15.72		0.85	
Blocks (B)	5	1.65	2.84*	0.45	<1.0	0.09	<1.0
A × B	10	1.26	2.17*	0.66	<1.0	0.17	1.55
*S*s in AB	75	0.58		0.78		0.11	

* $p < .05$.
*** $p < .001$.

of CRs (Table 1). The mean CRs per block were: F = 4.4; C = 5.6; and I = 5.3. A significant blocks effect indicated that the number of CRs varied from block to block, and a significant Blocks×Groups interaction suggested that this variation differed among groups. These data were further examined by a trend analysis (Edwards, 1960). Only Group F showed a significant linear trend, $F(1, 75) = 13.79$, $p < .001$, decreasing in CRs over blocks. The other groups failed to show a linear component different from a zero slope, $F(1, 75) = 2.53$ and <1.0 for Groups I and C, respectively. Thus, only the facilitating instructions affected CRs over blocks.

The analysis of the frequency of total SRs showed a significant difference among groups (Table 1). The facilitating instructions resulted in the highest rate of SRs (*M* over six blocks = 52.5), while the control group yielded a lower rate (*M* = 36.7), and inhibiting instructions gave the lowest rate (*M* = 26.0). A comparison of these effects with the number of CRs showed that the frequency of SRs in the control group approximated closely the actual number of CRs given, while Group F gave considerably more SRs than CRs and Group I gave fewer SRs than CRs (Fig. 2). Thus the instructions resulted in a shift of *S*'s tendency to maintain a congruence between the number of SRs and CRs.

An analysis on the arc-sine transformations of the proportion of incorrect SRs supported the preceding findings indicating significantly less accuracy in administrating SRs in the group receiving facilitating instructions than in the other groups. While Group F gave a proportion of incorrect SRs ranging between 43% and 60%, the C and I groups ranged between 18% and 24%. This indicated that the considerably higher number of SRs in Group F also included a higher proportion of incorrect SRs. It is interesting to note that the instructions affected SRs but did not change response accuracy in Groups C and I. Only in Group F was there a significant decrease in CRs over blocks. These findings suggest that accuracy of responding and probability of rewarding oneself for a response can be independently manipulated under some conditions.

EXPERIMENT III

The purpose of this experiment was to examine the generality of sr behavior, (*a*) when the stimulus properties of the discrimination task were changed from original learning to the test phase, and (*b*) when the original reinforcing stimulus could no longer be procured by the instrumental SR act. After learning the original discrimination task, any change in the four stimulus syllables will prevent *S* from knowing which is "correct." If we assume that SRs relate to some response by *S* during training which evaluates his performance, e.g., a subvocal statement "I was right," then a change in the syllables during the test phase should lower frequency of SRs, indicating the extent of generalization of sr to slightly different discriminative stimuli.

Omission of the external cue (light) in the test phase should decrease frequency and accuracy of SRs, if the SR serves mainly a substitute function for a previous externally controlled reinforcing stimulus and is learned by association with the reinforcing event (light).

Method

Subjects.—Forty-four male college students served as *S*s. An equal number of *S*s was assigned to each of four groups.

Apparatus.—The apparatus was the same as in Exp. I and II, except that a knife switch was added to disconnect the microswitch from the signal light on *S*'s side of the board.

Procedure.—The general procedure was identical to that used in Exp. I and II, except that all *S*s were run to a criterion of 7/10 correct responses. Ten blocks of 10 trials each were given in Phase II. The groups did not differ in trials to criterion nor in frequency of CRs on the last block of the training phase ($Fs < 1.0$).

Experimental design.—A 2 × 2 × 10 factorial design was employed. Half the *S*s were shown different sets of stimulus cards during the test phase. The last letter of each three-letter syllable seen during Phase I was changed during Phase II, without changing the association value of the syllables. The absence or presence of the original reinforcing stimulus (the green light) following *S*'s button pressing represented the second factor in the design. For half of the *S*s the green light was disconnected during Phase II. The instructions for *S*s in this group were the same as those used in Exp. I except for a minor modification to suit the experimental procedure. The *E* instructed *S* to "take this switch which I used to turn on the green light." For the remaining *S*s the signal light went on after they pressed the button in Phase II, as in the earlier experiments.

Results

The mean frequencies of CRs over blocks are shown in Fig. 3. When the list was changed a CR was defined as the syllable previously correct which now had one different letter. Both groups under changed-list conditions showed a drop in CRs at the outset of Phase II but the same frequency of CRs was maintained over blocks. For these groups the light resulted in slightly more accurate performance. Groups receiving the same list in Phases I and II showed a slightly better performance when the light did not flash. The latter groups yielded a relatively stable performance, with accuracy approximately equal to the number of CRs required for criterion learning (7/10). An analysis of variance on these data (Table 2) indicates, as expected, that changing the originally learned syllables in Phase II resulted in significantly fewer CRs. The absence of the original reinforcing light produced no

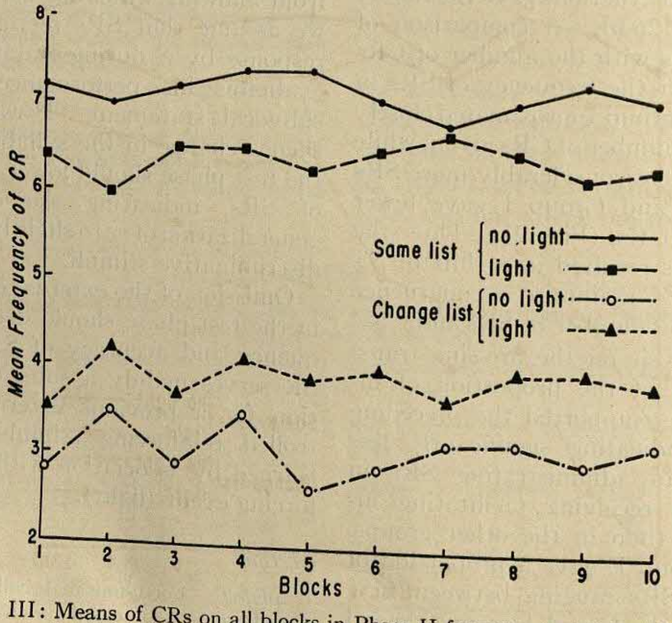


FIG. 3. Exp. III: Means of CRs on all blocks in Phase II for groups receiving the same or a different list than in Phase I, and a light or no light stimulus following SR.

TABLE 2

Analyses of Variance on SRs, CRs, and Arc-Sine Transforms of Proportion of Incorrect/Total SRs in Exp. III

Source	df	CR		SR		ISR/TSR	
		MS	*F*	*MS*	*F*	*MS*	*F*
Between *S*s							
List (A)	1	1174.91	98.07***	682.51	18.46***	63.89	27.89***
Light (B)	1	0.06	<1.0	64.15	1.74	8.38	3.66
A × B	1	68.02	5.68*	7.64	<1.0	11.35	4.95*
*S*s in groups	40	11.98		36.97		2.29	
Within *S*s							
Blocks (C)	9	0.47	<1.0	21.91	9.17***	0.52	2.60*
A × C	9	0.84	1.24	12.87	5.39***	0.33	1.65
B × C	9	0.21	<1.0	1.05	<1.0	0.27	1.35
A × B × C	9	0.53	<1.0	0.46	<1.0	0.19	<1.0
*S*s in blocks × groups	360	0.68		2.39		0.20	

* $p < .05$.
*** $p < .001$.

changes in CRs, when all groups were combined. However the differential effect of the light for list-same and list-changed conditions seen in Fig. 3 is reflected in the significant List × Light interaction (Table 2). The presence of the light thus enhanced correct responding only when other elements of the original learning situation were changed.

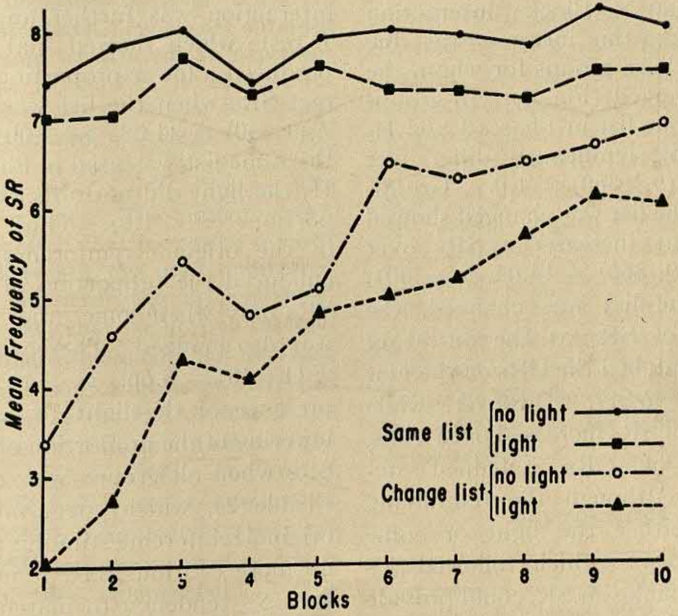


Fig. 4. Exp. III: Means of SRs on all blocks in Phase II for groups receiving the same or a different list than in Phase I, and a light or no light stimulus following SR.

Figure 4 presents the mean frequencies of SRs during Phase II. When the list was changed, SRs were low at first and gradually increased until they approached the frequencies observed in the groups with the same lists. The presence of the light on SRs had a somewhat different effect from that observed on CRs. Absence of the light tended to yield higher SR curves under both list conditions.

Analysis of the frequency of SRs revealed a significantly lower number of SRs when the list was changed (see Table 2). Using the same list, approximately the same number of SRs was obtained as the number of correct responses required for criterion learning (7/10). Mean SRs per block ranged from 7 to 8.2 in the two groups. When the list was changed however, SRs were lowered to a range of 2.0 to 6.8. A significant main blocks effect for all groups suggested that SRs increased over blocks. However, a breakdown of the significant List × Blocks interaction revealed that this increase was due mainly to those groups for whom the list was changed (Fig. 4). In groups with the same list in Phases I and II, SRs did not significantly differ over blocks, $F(9, 360) < 1.0$. Groups for whom the list was changed showed a significant increase in SRs over blocks, $F(9, 360) = 14.04$, $p < .001$.

When the list was changed, the frequency of SRs at the outset of Phase II matched the CRs on the new list fairly closely. However, while frequency of SRs increased over trials, frequency of CRs remained unchanged. Although the reinforcing stimulus, either the light or some covert response which followed *S*'s decision to make an SR, could provide no further information about response accuracy, *S*s showed increasing confidence in their responses. In the groups with the same list the SR response was at an asymptotic level (approximately 7.5/10) at the outset of Phase II and it did not increase further over blocks (Fig. 4). At the end of 10 blocks during Phase II, *S*s who were exposed to the changed lists approached the frequency of SRs yielded by the groups with the same list. Since CRs did not change, *S*s gave more SRs for incorrect responses as Phase II progressed. As shown by the *F* for Light and the Light × Block interaction (Table 2), the presence of the light did not significantly affect the number of SRs.

The analysis of variance on arc-sine transformations on the proportion of incorrect SRs showed that the change in the list resulted in significantly higher proportions of incorrect SRs. Further, this proportion increased significantly over blocks for the total sample (Table 2). The List × Light interaction was further analyzed by *F* tests which showed that the light produced a lower proportion of incorrect SRs when the list was changed, $F(1, 40) = 24.08$, $p < .001$. When the same list was used in Phases I and II, the light did not affect these proportions ($F < 1.0$). Thus, a change in the original reinforcing stimulus enhanced the proportion of incorrect SRs only when some other condition was also changed. There was a trend, $F(1, 40) = 3.66$, $p < .10$, for the presence of the light to result in a lowering of the proportion of incorrect SRs when all groups were combined (Table 2), which suggests that the original reinforcing stimulus may make *S*s more cautious, i.e., it may maintain *S*'s tendency to match *E*'s final rate of reinforcement in Phase I.

rather than simply increasing indiscriminate administration of SRs.

DISCUSSION

The results of all three experiments indicate that the behavior of *S*s who take over *E*'s reinforcing operations is systematically related to level of original learning, instructions about criteria for SRs, and similarity of the original and test lists. The *S*'s decision to make an SR is contingent, by instructions, on his own response to emission of the syllables (e.g., "I was right," "that's good," etc.). This judgment and other individual parameters which may determine the rate at which he judges his behavior to be correct control the probability of occurrence of SRs and were affected by the independent variables used in this study.

The decision toward sr is similar to behavior in psychophysical experiments in which a Yes or No response is required. Therefore, portions of the signal detection model (Goldiamond, 1962; Tanner & Swets, 1954) may be applicable. In both situations receiver operating characteristic (ROC) curves are obtained for yes-saying when a signal is present (a response is correct), or absent (incorrect). Variables affecting SR should then be examined with regard to their consequence on total SRs (yeses) and also on the number of incorrect SRs (false alarm rate). The present experiments suggested that increasing original learning increased the total number of SRs but decreased the proportion of incorrect ones; that facilitating instructions resulted in more SRs but also in a higher proportion of incorrect SRs; and that changes in the list decreased the total number of SRs but increased the proportion of incorrect SRs. Application of this model suggests a number of variables which may modify the total number of SRs and the ratios of incorrect to correct SRs independently. One important group of such variables would be the characteristics of individual *S*s (personality, attitude, etc.).

The reinforcing properties of SRs have not been directly studied here. Specifically, it remains to be seen whether an SR alters the probability of the preceding response on subsequent trials. In the present studies it was shown that *S*'s tendencies to judge responses to be correct can be altered. Whether this change also affects accuracy of the preceding response is a separate question. Some indications that SR *may* have reinforcing properties come from the findings that in all three studies, under different conditions, the level of response accuracy is maintained over as many as 100 trials without a significant decline (extinction), except for Group F in Exp. II. In this group the decrease could be attributed to the increasing number of SRs which, in turn, must have resulted in frequent SRs following incorrect responses. Thus incorrect responses may have been reinforced often. In Exp. III, groups for whom the lists were changed showed a parallel effect. In the no light group, more SRs and more CRs were found than in the light group. In these groups which had the difficult task of generalizing from an original list, the SRs may have served also to reinforce incorrect responses. Finally, in Exp. I, the greater uncertainty of Ss in the LL group is coupled with a relatively higher number of incorrect SRs than in any other group. Again, excess SRs could have reinforced both correct and incorrect responses, once they were emitted in Phase II. Since the present results suggest that maintenance and extinction of human behavior may be directly affected by SR, the relationship between SRs and CRs appears to be worthy of further study.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

GOLDIAMOND, I. Perception. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology.* New York: Basic Books, 1962. Pp. 280–340.

KANFER, F. H., BRADLEY, M. M., & MARSTON, A. R. Self-reinforcement as a function of degree of learning. *Psychol. Rep.*, 1962, 10, 885–886.

MARSTON, A. R., & KANFER, F. H. Human reinforcement: Experimenter and subject controlled. *J. exp. Psychol.*, 1963, 66, 91–94.

TANNER, W. P., JR., & SWETS, J. A. A decision-making theory of visual detection. *Psychol. Rev.*, 1954, 61, 401–409.

(Received September 17, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 255–259

REVERSAL PERFORMANCE BY RATS FOLLOWING OVERLEARNING WITH AND WITHOUT IRRELEVANT STIMULI[1]

KEITH N. CLAYTON

Vanderbilt University

In 2 separate experiments, degree of learning (moderate vs. overlearning) was manipulated simultaneously with irrelevant cues conditions (present vs. absent) during discrimination learning and reversal by rats. Massed position discrimination training was given to criterion or to 125–150 trials beyond criterion at which time reversal began. Reversal also was taken to criterion. In direct contrast to earlier studies, overlearning significantly retarded reversal performance in both experiments. The results also suggest that reversal performance following overlearning is made even more difficult by the presence of irrelevant stimuli. The fact that some studies have found that overlearning facilitates reversal and other studies have not, therefore, apparently is not due solely to the fact that the successes included irrelevant cues while the failures did not.

The present study was stimulated by the conflicting evidence with respect to the effect of overlearning a discrimination on its subsequent reversal. Beginning with Reid (1953) several *E*s have shown that overlearning facilitates reversal. Recently, however, Hill, Spear, and Clayton (1962) reported a series of three experiments in which all failed to show a significant reversal superiority of overlearned groups. Similar findings involving four separate experiments have been reported by D'Amato and Jagoda (1962).

While all the studies differ on numerous procedural details, the present study investigated the hypothesis that the critical difference between the successes and failures was whether or not irrelevant cues were present during acquisition and reversal. In all cases, the studies showing a facilitory effect of overlearning trained rats on a discrimination in which irrelevant cues were present. Either rats were trained on a nonspatial discrimination in which position was irrelevant (Place Groups, Brookshire, Warren, & Ball, 1961; Capaldi & Stevenson, 1957; D'Amato & Jagoda, 1961; North & Clayton, 1959; Exp. I, Pubols, 1956; Reid, 1953) or were trained on a position discrimination with irrelevant brightness cues (Response Groups, Brookshire et al., 1961; Exp. II, Pubols, 1956). The studies not showing the effect, however, trained rats on a position discrimination without irrelevant brightness or extramaze cues. The suggestion, then, is that the presence of irrelevant cues is essential to the overlearning reversal effect (ORE).

This empirical observation is not without its theoretical counterpart. Reid (1953) and Pubols (1956), for example, have suggested that the ORE is due to the acquisition, during overlearning, of an observing response which serves to mediate and improve subsequent reversal performance. Assume the observing response which is

---

[1] This study was supported in part by Grant M-5576 from the National Institute of Mental Health. The author wishes to thank William B. Harrison for assistance in running *S*s.

acquired, or refined, during overlearning is that of ignoring irrelevant stimulation. Reversal performance, therefore, would presumably be facilitated since learning during reversal would occur primarily with respect to relevant discriminanda. If irrelevant cues were absent, or minimized, however, *S* would not be expected to profit from the overlearning experience.

A factorial design was utilized in two experiments to test the hypothesis. Degree of learning (moderate vs. overlearning) was simultaneously manipulated with irrelevant-cues conditions (irrelevant cues present vs. no irrelevant cues present).

EXPERIMENT I

Method

Subjects and apparatus.—The *S*s were 20 experimentally naive male albino rats of the Sprague-Dawley strain approximately 4 mo. of age. They were randomly assigned to each of the four conditions (5 per cell).

The apparatus was a gray enclosed **T** maze 3 in. wide and 7 in. deep throughout. The lengths of the start box, stem, and arms were 8½ in., 14½ in., and 9¾ in., respectively. Aluminum one-way swinging doors separated the start box from the stem (operated by *E*) and the arms from the goal boxes (operated by *S*). The reward was located in a circular hole in the floor of the goal box just behind the goal-box door. This door was locked on the nonrewarded side. Sliding Plexiglas guillotine doors, located just beyond the choice point, were used for forcing purposes (during pretraining only) and to prevent retracing. The entire apparatus was covered by Plexiglas.

The discriminanda defining the irrelevant-cues condition were: (*a*) the pattern on the goal-box doors, (*b*) a striped plaque on the floor of the choice point, and (*c*) the location of black curtains hung on the otherwise white walls of the experimental cubicle. The two goal-box doors, which were interchangeable, were painted with ¾-in. wide black and white stripes. One door had vertical stripes; the other, horizontal. The choice-point plaque measured 2½ × 14 in. and extended 7 in. into each arm from the center of the choice point. Half the plaque was painted with ¾-in. wide black and white stripes running in the direction of the length of the plaque and on the other half the stripes were perpendicular to the length of the plaque. The plaque, therefore, could be placed in the maze so that either horizontal or vertical stripes led into one arm. The curtains, extending from the top to 3 ft. above the maze, were placed in either of two positions: either covering the entire wall along the "top" of the **T** maze approximately 1 in. from the maze, or covering the two walls beyond the goal boxes, approximately 2 ft. behind the goal-box doors. For the Irrelevant-Cues group, therefore, these discriminanda were randomly and independently changed from trial to trial, whereas for the No-Irrelevant-Cues group their position never changed.

Prehandling.—For 7 days prior to the beginning of pretraining each *S* was given prehandling daily. During these 2-min. prehandling sessions *S* was permitted to explore a table top, given access to four of the food pellets later to be used as reward, and picked up and replaced at least five times by *E*.

Pretraining.—Pretraining consisted of 4 days training in the **T** maze. On Day 1, *S* was given one free, rewarded, trial in which the goal-box door was manipulated when necessary by *E*. On Day 2, *S* was forced to and rewarded on the side opposite that chosen on Day 1. On each of Days 3 and 4, *S* was given four rewarded trials, forced only as required to maintain equality of experience with the two sides. Throughout pretraining no irrelevant discriminanda were present. The reward during pretraining and thereafter was two 45-mg. Noyes food pellets.

Discrimination training.—Five trials on a position discrimination were given daily, with the exception of the first day on which four trials were given. The positive side for each *S* was the nonpreferred side as determined by the free trial pretraining choices. Four to 7 min. elapsed between the trials. The daily ration of 12 gm. of powdered food was given 20–30 min. following the daily session.

The criterion of learning was two or less errors in 3 consecutive days with the additional restriction that the last day be errorless. The moderate learning *S*s were reversed on the day following reaching the criterion, the overlearning *S*s received an additional 25 days (125 trials) of acquisition. The only procedural change during reversal was the change in the locus of the food reward. Reversal training continued until the criterion of learning was again reached.

Results

Acquisition.—The frequency of errors and number of days to criterion were tabulated for each *S*. By analyses of variance of both sets of data no source, irrelevant cues, degree of learning (a test of comparability) nor interaction, was statistically significant. The Irrelevant-Cues group made more errors (a mean of 8.0 vs. 5.6 for the No-Irrelevant-Cues group) but the $F\,(1,\,16) = 2.37$ fell short of significance. The Overlearning group made a mean of 1.4 errors during the 125 trials beyond criterion.

Reversal.—A trend analysis of the reversal data included as sources: degrees of learning, irrelevant-cues condition, days (5), and all interactions. Only the *F*s for days and degrees of learning were significant. For days, $F\,(4,\,64) = 54.82$, $p < .01$, and for degrees of learning, $F\,(1,\,16) = 7.30$, $p < .05$. The overlearned *S*s made more errors during reversal than the moderately learned *S*s (means of 14.7 and 10.8, respectively). The triple interaction was in the direction exactly opposite that hypothesized, i.e., the inferiority of the overlearned *S*s was more pronounced with irrelevant cues. This interaction was not significant, however; $F\,(4,\,64) = 2.09$, $p < .10$. The overlearned *S*s also took significantly more days to reach the reversal criterion [a mean of 7.40 vs. 5.80 for the moderate *S*s, $F\,(1,\,16) = 5.17$, $p < .05$].

EXPERIMENT II

In Exp. I overlearning retarded reversal learning regardless of the presence or absence of irrelevant cues. However, in order to test adequately the hypothesis that the presence of irrelevant stimuli affect the ORE it seems reasonable to demand independent evidence that the irrelevant cues condition was in fact manipulated. In no analysis did irrelevant cues significantly retard position discrimination despite evidence to the contrary (e.g., Restle, 1957). Perhaps the discriminanda were not *functionally* irrelevant. The use of irrelevant patterns inside the maze and varying spatial cues outside the maze may not be as potent as varying black and white inside the maze, as was done in the previous overlearning studies (Brookshire et al., 1961; Pubols, 1956). For these reasons Exp. I was repeated with irrelevant intramaze black and white stimuli.

Method

Subjects and apparatus.—The *S*s were 20 experimentally naive male albino rats of the Sprague-Dawley strain 112–148 days of age at the beginning of acquisition.

The apparatus was a discrimination box composed of three parts: start, choice, and goal chambers. The start box was 3¾ × 9 in. and separated from the choice area by a vertical sliding wooden guillotine door. The choice chamber was shaped as a truncated triangle with the following dimensions: top, 3¾ in.; sides, 6 in.; and base, 7½ in. The start and choice areas were painted gray and the entire apparatus was 6 in. deep and covered by hardware cloth. Two wooden one-way swinging doors led from the choice area into two adjacent goal boxes which were 4 in. wide and 12 in. long. The goal area was the last 3 in. of the goal box and was concealed from the doors by a 2-in. high metal partition. Reward was placed in a small plastic caster in this goal area.

The irrelevant-cues conditions were defined by the brightness values of the goal-box doors and goal boxes. These goal boxes were interchangeable so that the following conditions could be used: gray doors and boxes on both sides, white door and goal box on the left with black door and box on right, or black door and box on the left with white door and box on right. For the No-Irrelevant-Cues group the gray boxes were used, for the Irrelevant-Cues group the black and white cues were altered randomly with respect to position.

Prehandling and pretraining.—Prehandling, as described in Exp. I, was administered for 8 days. Pretraining was equivalent to that in Exp. I except that 10 trials were given in 3 days instead of 4. All *S*s were pretrained with gray doors and goal boxes.

Discrimination training.—Experiment II training differed from Exp. I on the following: 10 trials were given daily, approximately 15 sec. elapsed between the trials, the criterion of learning was 18 or more correct responses on 2 consecutive days—with no more than one error on the last day, and overlearning consisted of 15 days (150 trials). Incorrect responses led to an empty goal box and the one-way swinging door prevented retracing.

Results

Acquisition.—As in Exp. I, analyses of variance of the acquisition errors and days to criterion data revealed no significant effects. Overall, the *S*s required means of 9.7 errors and 4.1 days to reach criterion. The overlearning *S*s made a mean of 1.8 errors during overlearning.

Reversal.—An analysis of variance of the days to criterion data revealed only a significant effect of the irrelevant-cues condition, $F\ (1,\ 16) = 9.19$, $p < .01$. The irrelevant-cues *S*s took a mean of 8.2 days vs. 4.9 days for the no-irrelevant cues *S*s. No other sources were significant by this analysis. However, a trend analysis of the errors data revealed significant main effects for all three main sources and the triple interaction. For degrees of learning and irrelevant-cues condition the Fs $(1,\ 16) = 4.66$ and 15.82, respectively, $p < .01$. For the triple interaction, $F\ (4,\ 64) = 9.72$, $p < .05$. Figure 1 depicts the interaction, which may be described as follows: the inferiority of the overlearned *S*s was accentuated when irrelevant cues were present. This interaction, it will be recalled, was in the same direction but not significant in Exp. I.

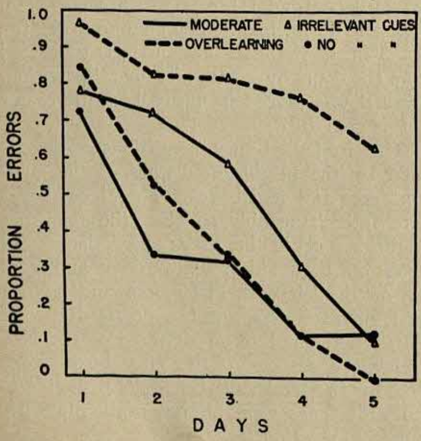


FIG. 1. Choice performance during the first 5 days of reversal (Exp. II).

DISCUSSION

The hypothesis that the conflicting evidence is due to whether or not irrelevant stimuli are present was not supported by the present study. Rather than finding the ORE only when irrelevant cues are present, the results show that the ORE does not occur when functionally irrelevant cues are present. The critical difference between the successes and failures remains to be found. Together, the recent studies have begun to rule out several singular factors, perhaps the difference is a combination of factors.

The present finding that reversal of a position discrimination was *more* likely to be impaired by irrelevant stimuli following overlearning was unexpected from the hypothesized observing response mechanism. Similar notions, e.g., those of Restle (1957) that irrelevant cues are neutralized, and of Lawrence (see Goodwin & Lawrence, 1955) that irrelevant cues become more distinctive, would also have difficulty accounting for this finding. A primary need, apparently, is a clear specification of the conditions under which the presumed mediator transfers from acquisition to reversal.

REFERENCES

BROOKSHIRE, K. H., WARREN, J. M., & BALL, G. G. Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.*, 1961, **54**, 98–102.

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

D'Amato, M. R., & Jagoda, H. Analysis of the role of overlearning in discrimination reversal. *J. exp. Psychol.*, 1961, **61**, 45–50.

D'Amato, M. R., & Jagoda, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, **64**, 117–122.

Goodwin, W. R., & Lawrence, D. H. The functional independence of two discrimination habits associated with a constant stimulus situation. *J. comp. physiol. Psychol.*, 1955, **48**, 437–443.

Hill, W. F., Spear, N. E., & Clayton, K. N. T maze reversal learning after several different overtraining procedures. *J. exp. Psychol.*, 1962, **64**, 533–540.

North, A. J., & Clayton, K. N. Irrelevant stimuli and degree of learning in discrimination learning and reversal. *Psychol. Rep.*, 1959, **5**, 405–408.

Pubols, B. H., Jr. The facilitation of visual and spatial discrimination reversal by overtraining. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

Restle, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" controversy. *Psychol. Rev.*, 1957, **64**, 217–228.

(Received September 19, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 260–263

EFFECTS OF DELAY ON SUBSEQUENT RUNNING UNDER IMMEDIATE REINFORCEMENT[1]

JOSEPH A. SGRO[2] AND SOLOMON WEINSTOCK[3]

Lehigh University

3 groups of 25 albino rats were given 36 training trials in a runway with 0, 7.5, and 15 sec. delay of reinforcement, respectively. All Ss were then given 36 trials with immediate reinforcement followed by 18 extinction trials. All trials were administered under 21–22 hr. of water deprivation and with a 24-hr. intertrial interval. In the delay phase Group 0 ran significantly faster than Group 7.5 and Group 15. In the no-delay phase, Group 15 ran significantly faster than Group 0 and Group 7.5. Covariance analysis, in which the terminal speed of running in the no-delay phase was the control variable, showed that the effect of delay was to increase resistance to extinction.

The suggestion has been made that nonreinforcement may be treated theoretically as a special case of delay of reinforcement, that is, as a very long or "infinite" delay. On this basis, one may expect to find similar effects resulting from both variables. Studies on partial delay of reinforcement (e.g., Crum, Brown, & Bitterman, 1951; Scott & Wike, 1956) have shown that *S*s who receive delay on some proportion of their training trials are more resistant to extinction than no-delay *S*s.

Similarly, a number of studies (e.g., Jensen & Cotton, 1960; Lauer & Carterette, 1957; Lauer & Estes, 1955) have shown that extinction trials lead to more rapid running in a subsequent conditioning session. The present study is designed to detect a comparable effect due to delay of reinforcement. To this end, three groups of rats were given training in a runway under 0, 7.5, and 15 sec. delay, respectively, followed by trials under immediate reinforcement. It was hypothesized that in the second phase of the experiment the delay groups would run more rapidly than the no-delay group.

METHOD

Subjects.—The *S*s were 75 female Wistar albino rats, 90–100 days old, at the beginning of the experiment. They were selected from a larger group of 90 *S*s. Each *S* received six trials with no delay of reinforcement and the 15 slowest *S*s (smallest mean reciprocal running times) were eliminated. Five *S*s were discarded during the study. One *S* bit *E* repeatedly, another failed to run on the early acquisition trials, and three died from unknown causes.

Apparatus.—A straight runway with an L shaped goal box was used. It was made of ¾-in. plywood and was 5 in. wide and 5 in. high throughout. The start box, runway, and goal box were 9½, 39, and 17 in. in length, respectively. Guillotine doors were placed at the beginning and end of the runway. The arm of the L shaped goal box was 12½ in. long.

A drinking tube, similar to tubes used during the regular watering periods, was located at the end wall of the L. A horizontal sliding door, 1½ in. in front of the end wall, prevented access to the drinking tube during the delay intervals and during extinction.

The entire apparatus was left in its natural finish and was covered by ¼-in. hinge-mounted Plexiglas. Lighting was provided by four 40-w. incandescent bulbs, spaced 21½ in. apart and mounted 42½ in. above a 31-in. high table which held the apparatus.

Slots were cut in the side walls 1⅛ in. from

---

[1] This study was conducted as the master's thesis of the first author under the direction of the second author. The latter is primarily responsible for the preparation of this report.

[2] Now at Texas Christian University.

[3] Now at Brooklyn College.

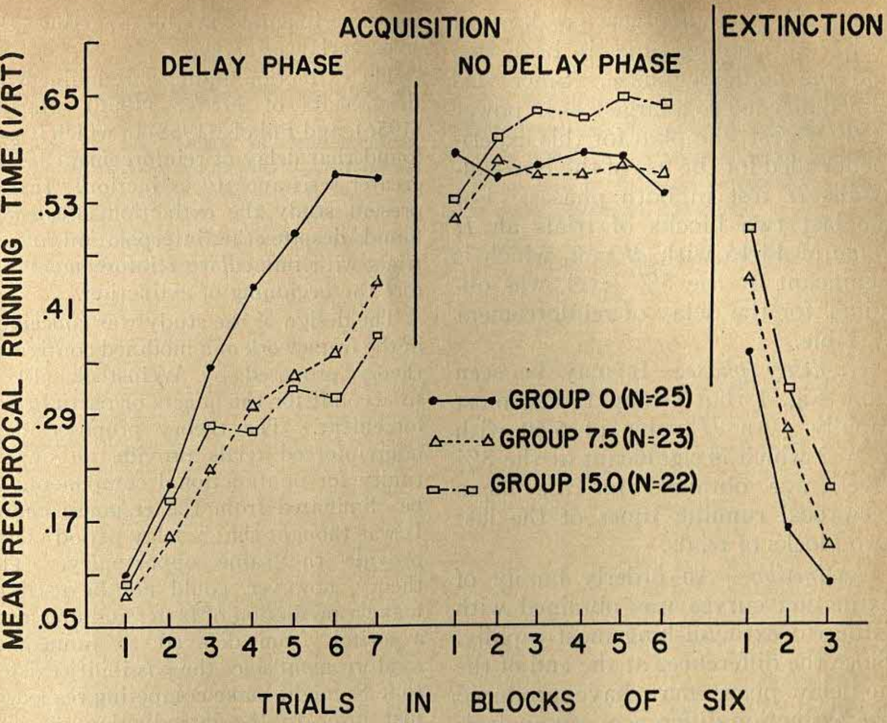


FIG. 1. Mean reciprocal running time plotted in blocks of six trials.

the start box and 8½ in. from the end of the runway to permit the passage of photocell beams. Two Standard Electric .01-sec. timers were started by a microswitch on the start box and were stopped by interruption of the photocell beams. Interruption of the beam in the goal box started a third timer which was used to time the delay period and the interval during which *S* was given access to the drinking tube.

Procedure.—The *S*s were assigned haphazardly to cages, three or four to a cage. After 2 days of habituation to the colony room, they were placed on a water deprivation schedule and allowed access to water for 1½ hr. each day. For the next 5 days, each *S* was handled approximately 1 min. daily.

Training trials were given under 21–22 hr. of water deprivation and with an intertrial interval of 24 hr. All *S*s were given six trials with no delay of reinforcement and were then assigned randomly to groups receiving 0, 7.5, and 15 sec. delay of reinforcement, respectively. On Trials 7 through 41 *S*s were reinforced according to the condition to which they had been assigned. On Trials 42 through 77, all *S*s received no delay of reinforcement. On Trials 78 through 96 *S*s received extinction.

The typical procedure on a trial consisted in *E*'s removing *S* from its home cage and placing it in the start box facing the guillotine door. Five seconds later *E* opened the door. After *S* interrupted the second photocell beam *E* lowered the goal-box door. On all reinforced trials *S* was allowed 30 sec. access to the drinking tube. On all extinction trials *S* was retained for 20 sec. in the goal box. The *S* was then removed from the apparatus and returned to its home cage.

RESULTS

Latency and running time measures for each trial were converted to reciprocals. All analyses were performed using the reciprocal measures.

Delay phase.—The mean reciprocal running times are presented in Fig. 1 in blocks of six trials. In a pilot study it was found that a small number of

*S*s in each group had very small mean reciprocal running times in both this and the no-delay phase. Since this bimodality led to a large loss in power of the *F* test, the plan for this experiment called for the use of the Kruskal-Wallis *H* test in both phases. For the last two blocks of trials an *H* value of 14.16 with $df = 2$, which is significant at the 5% level, was obtained for the delay of reinforcement variable.

No-delay phase.—It may be seen from Fig. 1 that Group 15 ran most rapidly. An *H* value of 6.60 with $df = 2$ which is significant at the 5% level, was obtained for the mean reciprocal running times of the last two blocks of trials.

Extinction.—An orderly family of extinction curves was obtained with Group 0 extinguishing most rapidly. Since the differences at the end of the no-delay phase may have produced the extinction differences, an analysis of covariance was performed on the mean reciprocal running time for the three blocks of extinction using the last two blocks of no-delay as the control variable, $F(2, 66) = 5.72$, $p < .05$.

Discussion

Spence (1956) cites a number of studies involving a switch from delayed to immediate reinforcement which did not show faster running after delay. It is difficult to decide whether differences in procedural and situational variables or lack of sensitivity in the earlier studies are responsible for the differences in the results. The present study was designed to detect a difference as small as 10% in the means of the extreme groups. A pilot study, which yielded somewhat larger mean differences and an estimate of the error variance, was used to choose the number of *S*s in each group. In addition, the 15 slowest running *S*s were eliminated prior to the introduction of the experimental variable to reduce the error variance further.

The extinction findings are in line with the results of Fehrer (1956), Harker (1956), and Pubols (1958) in which it was found that delay of reinforcement led to greater resistance to extinction. In the present study the extinction effect was found despite the interpolation of 36 trials with immediate reinforcement before the beginning of extinction.

The design of the study was conceived in the framework of a modified contiguity theory proposed by Weinstock (1958) to account for the effects of partial reinforcement. The theory proposed that nonreinforced trials provide an opportunity for nonfunctional components to be eliminated from the response chain. It was thought that a delay period might provide the same opportunity. The theory, however, could not be used to make a prediction of faster running after a switch from delayed to immediate reinforcement since the possibility existed that *S*s might make competing responses just prior to the introduction of reinforcement on the delay trials. The value of the theory was programmatic in that it suggested the possibility of the effect. However, no specific support for the theory can be claimed.

REFERENCES

Crum, J., Brown, W. L., & Bitterman, M. E. The effect of partial and delayed reinforcement on resistance to extinction. *Amer. J. Psychol.*, 1951, **64**, 228–237.

Fehrer, E. Effects of amount of reinforcement and of pre- and postreinforcement delays on learning and extinction. *J. exp. Psychol.*, 1956, **52**, 167–176.

Harker, G. S. Delay of reward and performance of an instrumental response. *J. exp. Psychol.*, 1956, **61**, 303–310.

Jensen, G. D., & Cotton, J. W. Successive acquisitions and extinctions as related to percentage of reinforcement. *J. exp. Psychol.*, 1960, **60**, 41–49.

Lauer, D. W., & Carterette, T. S. Changes in response measures over repeated acquisitions and extinctions of a running habit. *J. comp. physiol. Psychol.*, 1957, **50**, 334–338.

LAUER, D. W., & ESTES, W. K. Successive acquisitions and extinctions of a jumping habit in relation to schedule of reinforcement. *J. comp. physiol. Psychol.*, 1955, **48**, 8–13.

PUBOLS, B. H. Delay of reinforcement, response perseveration and discrimination reversal. *J. exp. Psychol.*, 1958, **56**, 32–40.

SCOTT, E. D., & WIKE, E. L. The effect of partially delayed reinforcement and trial-distribution on the extinction of an instrumental response. *Amer. J. Psychol.*, 1956, **69**, 264–268.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. *J. exp. Psychol.*, 1958, **56**, 151-158.

(Received September 19, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 264–273

LEARNING OF INTERVAL CONCEPTS: I. EFFECTS OF DIFFERENCES IN STIMULUS WEIGHTS[1]

CHARLES N. UHL[2]

University of Alberta

A type of learning task was studied in which a multiple-regression prediction equation defined the relationship between 3 stimuli and a criterion which the S's response attempted to predict. 7 tasks, each having a stimuli-criterion multiple R of unity, differed in the degree of disparity in the distribution of regression weights to stimuli from one extreme where all stimuli had equal weights to the opposite extreme where only 1 stimulus was weighted. In an 8th task none of the stimuli were weighted. 80 undergraduate Ss, 10 per task, were given 150 training trials. The functional relationship between performance and disparity of weights was nonmonotonic. Performance was poorest in tasks with small disparity of weights, slightly better with no disparity of weights, and markedly better with larger disparity of weights.

Concept formation studies traditionally have employed sets of stimuli whose attributes can be classified only in a nominal fashion, i.e., red or blue in color, triangular or square in form, etc. (e.g., Bourne, 1957; Hull, 1920; Reed, 1946). When stimuli exhibiting an ordered relationship between attributes have been employed, i.e., one or two forms in number, the ordering relation has not been a necessary part of the specification of the concept-to-be-attained. Responses made available to *S* in the traditional concept formation study also have been capable of nominal classification only, i.e., one or the other of two nonsense syllables, a left or a right button to press, etc.

Another characteristic of the traditional concept formation task is an all-or-none dependence between stimulus attributes and correct responses. For example, in a two-response situation the presence of one stimulus attribute (or one conjunction of attributes) always is the occasion for Response A_1 and the presence of a second attribute (or conjunction of attributes) always is the occasion for Response A_2. Degrees of uncertainty concerning the contingency between a stimulus dimension and the correct response rarely have been permitted.

The foregoing comments make it clear that the traditional study of concept learning has failed to deal with many significant features of human learning. Brunswik (1956) has demonstrated that relatively valid responses may depend reliably upon a number of stimulus dimensions each of which may have less than complete ecological validity (i.e., relevance). In the situations Brunswik investigated the attributes along a stimulus dimension exhibited an ordered relationship as did the response categories. Much behavior requires analysis which recognizes that stimulus and response values may be ordered, and that a response may be contingent simultaneously upon several stimulus dimensions. Adams (1953), following the classification of scales outlined by Stevens (1951), has proposed a classification of concepts as nominal, ordinal, interval, or ratio depending upon relations holding between values along the stimulus dimension(s) concerned. For present purposes, a concept is classified in terms of relations between response values as well as relations between stimulus values, i.e., an interval concept is one whose stimulus and response values can be located on an interval scale of measurement. The phrase "learning of interval concepts" refers to the formation of interval or ratio concepts which entail a weighted combination of stimulus dimensions in the selection of a response along a single response dimension.

---

[1] Supported in part by a grant (No. 4894) from the National Institute of Mental Health, United States Public Health Service, and in part by a grant from the General Research Fund of the University of Alberta.

[2] Now at the University of Utah.

The present study was designed as a test of a general method for the study of the learning of interval concepts. Several workers (Azuma, 1960; Smedslund, 1955; Summers, 1962; Uhl, 1960) recently have reported experimental investigation of the learning of interval concepts. The primary aim of these studies was to demonstrate the extent to which interval concepts can be learned. Summers found that *S*s on the average weighted cues approximately proportional to their validities. The correlation between responses and correct answers rose to +.40 and showed no sign of having reached a limit when training was stopped. None of these studies involved a direct comparison of the extent of learning under different task conditions. The present study seeks to determine the effects of the distribution of stimulus weights upon learning. The learning of interval concepts as conceived of here consists, in part, of a discrimination of the degree of contribution of a stimulus dimension in the determination of the criterion, i.e., stimulus regression weight. A given stimulus dimension may be assigned a weight, in combination with the other stimulus dimensions, ranging anywhere from 0% to 100% of the available regression weight. The prediction to be tested is an extrapolation from the results obtained in classical studies of the discrimination of stimuli varying along a single dimension. The ease of discrimination of a difference in the regression weight of two stimulus dimensions should be an increasing function of the magnitude of difference between their respective weights, provided that the size of the multiple correlation is held constant. Zero or near zero differences between stimulus weights need not be discriminated for accurate responding. Moderate weight differences will be difficult to discriminate and a failure of discrimination will impede response accuracy. Large weight differences will be discriminated readily and response accuracy will be high. The overall function relating response accuracy at the end of training to the magnitude of differences in stimulus weights will be bow shaped.

METHOD

Subjects.—The *S*s were 80 undergraduate students from introductory psychology classes at the University of Alberta. They were randomly assigned to eight groups of 10 *S*s each. Half of the *S*s in each group were male and half were female.

Apparatus.—Each *S* was seated at a table on which was located a stimulus panel and a response panel. The stimulus panel was located 5 ft. in front of *S* and consisted of three rows of nine 1-in. pilot lights. The top row of lights was red in color, the middle row was green, and the bottom row was amber. The lights within a row were 3 in. apart on center and the lights between rows were 4 in. apart on center. A buzzer also was concealed in the stimulus panel. The response panel was immediately in front of *S* and consisted of a row of nine ½-in. amber pilot lights at the top (the criterion lights), nine toggle switches in the middle row, and a row of nine ½-in. white pilot lights at the bottom (the response lights).

An automatic programing device controlled the sequence and temporal intervals of the

TABLE 1

Stimulus Validities Squared of Tasks

Task	Stimulus		
	1	2	3
1	.319	.333	.347
2	.212	.333	.454
3	.105	.333	.561
4	.000	.331	.668
5	.000	.224	.775
6	.000	.117	.882
7	.000	.010	.989

stimulus lights, the criterion lights, and the buzzer. The *S* responded by moving the handle of one of the toggle switches forward. Immediately following a response a criterion light was turned on to give *S* feedback information and the response light directly beneath the switch responded to was turned on to give *S* a check on the location of his response. The *S* was able to compare the location of the response light with the location of the criterion light.

The *E* was seated behind and to the left of *S*. A panel in front of *E* duplicated the criterion and response lights of *S*'s response panel for recording purposes.

Task.—On each trial throughout the experiment the stimulus display consisted of one light glowing in each row. The appearance of a stimulus display was similar to that of a profile of three test scores. A multiple-regression prediction equation was used to determine which of the criterion lights would glow in each stimulus display. That is, each row of stimulus lights was treated as a predictor variable with values of 1–9 from left to right. The row of criterion lights on the response panel was treated as the dependent variable in the prediction equation. Criterion lights from left to right corresponded to values of the dependent variable from 1 to 9. Calculated criterion values were rounded to the nearest integer.

Thus the task for *S* was to learn to predict which criterion light would appear given the displayed values on the three stimulus dimensions. In effect, *S* had to attempt to learn the regression weights and the constant entered in the multiple-regression prediction equation.

The stimulus values displayed were drawn from a table of random normal numbers prepared by Hoffman (1959) for which the columns yield near zero intercorrelations each with the other. Values of each stimulus dimension were drawn from separate columns of the table. The obtained product-moment intercorrelations of stimulus dimensions within a series of 50 observations did not exceed ±.10. Each consecutive sequence of 25 stimulus values was forced to yield a standard deviation of 2.0 ± 0.25. The sequence of stimulus values displayed was the same for all *S*s regardless of experimental condition. The standard deviation of the criterion dimension as substituted into the multiple-regression prediction equation was arbitrarily set at 2.0.

The seven learning tasks investigated differed with respect to the size of the regression weight assigned to a stimulus dimension. The seven tasks differed in degree of disparity between weights in a linear manner. Table 1 presents the squared validity coefficient of each stimulus dimension in Tasks 1–7. The squared validity coefficient gives the proportion of regression weight assigned to a stimulus. The multiple correlation of stimulus dimensions with the criterion dimension was set at 1.0 for Tasks 1–7 and at 0.0 for an eighth comparison task (to be referred to as T0).

Procedure.—All *S*s were tested individually in a 1-hr. session. Each *S* was given instructions which acquainted him with the operation of the apparatus. He was told to attend to the three rows of stimulus lights and to use the stimulus information in selecting a response, but he was given no hint that the solution was of the multiple-regression form or that the stimulus dimensions might differ in weight. Although the task was found to be intrinsically motivating, *S* was instructed to try to earn points in order to effect more standardized incentive motivation and to reinforce *S* for close, in addition to perfect, predictions. The number of points awarded depended upon the absolute discrepancy between the response and the criterion: 5 points—0 discrepancy, 3 points—discrepancy of 1, 1 point—discrepancy of 2, and 0 points—discrepancy of 3 or more. The *S* was informed that the maximum point total was 750 and that a total of 375 was considered a passing performance. The *S* was informed of his point total after every 50 trials.

Four practice trials were administered and *S* was invited to ask questions before proceeding. The training series proper consisted of 150 trials run continuously. A trial was initiated by the onset of the three stimulus lights. At 10 sec. the buzzer sounded for ½ sec. to warn *S* that time was running out. At 15 sec. the stimulus lights, the criterion light, and the response light were turned off.

The *S* was permitted to respond by throwing a toggle switch at any time during the presentation of the stimuli. The criterion and response lights came on immediately after the occurrence of a response and remained on until 15 sec. had elapsed. The intertrial interval was 5 sec.

RESULTS

Sex differences.—No significant differences between sexes in error scores or response validities were found at the .05 level. Therefore, sexes were combined for the analysis of all results which follows.

Response validity.—The correlation (Pearson r) of responses with criterion values is referred to as response validity. Response validity reflects the accuracy of stimulus weighting and combining uninfluenced by the variance or the mean of the response distribution. A response validity coefficient within successive blocks of 50 trials was obtained for each *S*. These response validities were transformed by Fisher's method for all statistical treatments. Figure 1 presents the mean transformed response validities over blocks of trials for all tasks. A one-way analysis of variance of trends was performed on the response validities and is summarized in Table 2. Task 0 was not included in the analysis of variance since it was not on the same dimension of treatment variation as the remaining tasks. Response validities for all tasks (except T0) increased over blocks of trials indicating that learning took place. It should be noted that response validity showed no increase in T0. Tasks differed significantly over the three blocks combined. The rank order by tasks of mean response validity from high to low was T7, T6, T5, T4, T3, T1, and T2. This indicates that the attainment of response validity was facilitated by a high degree of disparity between stimulus weights, with the exception of T2 for which the disparity was slight. The significant Blocks × Tasks interaction suggests that the rate of learning differed between tasks. However, this effect was almost entirely due to the more rapid rate of learning in T7. When a separate analysis of variance was conducted without T7 the interaction was no

TABLE 2

ANALYSIS OF VARIANCE OF TRENDS OF RESPONSE VALIDITIES

Source	*df*	*MS*	*F*
Tasks (T)	6	7.39	16.47***
Linear	1	82.15	182.56***
Quadratic	1	37.33	82.96***
Cubic	1	8.40	18.67***
Error (a)	63	0.45	
Blocks of trials (B)	2	5.92	59.20***
Linear	1	10.17	101.70***
Quadratic	1	1.68	16.80***
T × B	12	0.67	6.70***
Linear	6	6.31	63.10***
Quadratic	6	2.07	20.70***
Error (b)	126	.10	

*** $p < .001$.

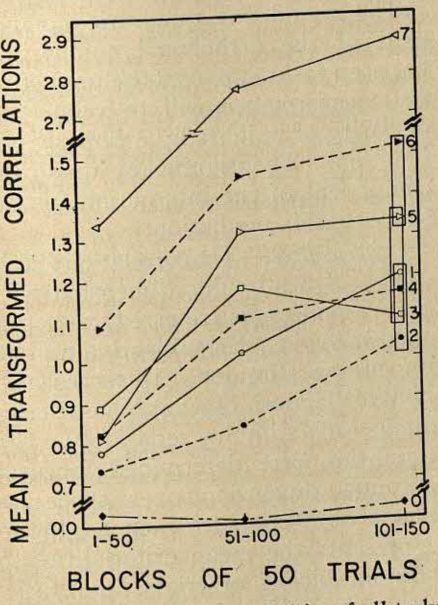


FIG. 1. Mean response validities of all tasks.

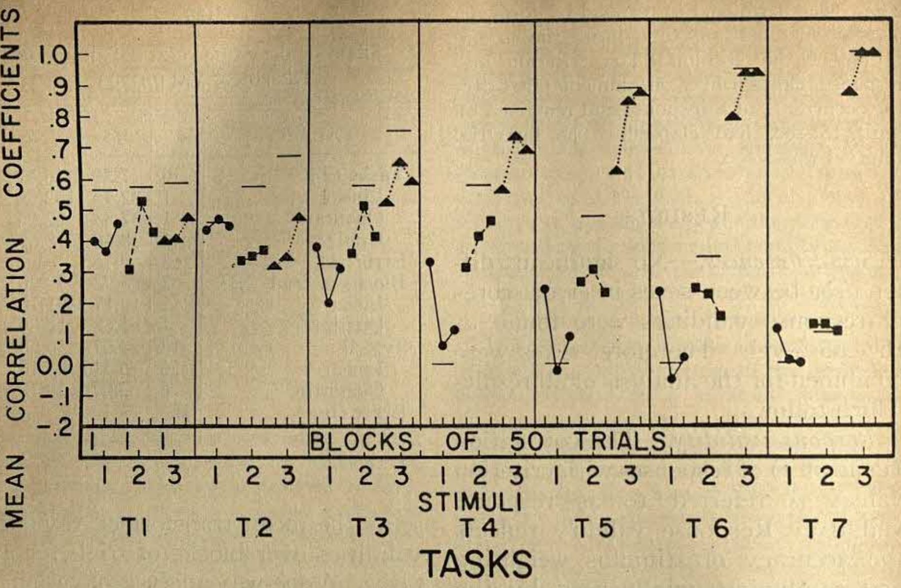


FIG. 2. Mean criterialities of Tasks 1–7. (Tasks are separated by vertical panels. Three connected points represent mean criterialities for a given stimulus over blocks of 50 trials. The horizontal line spanning a set of connected points represents the validity of the stimulus.)

longer significant, $F(10, 108) = 1.6$, $p > .05$.

The terminal level of response validity is of particular interest in connection with the effects of disparity of stimulus weighting. All tasks (except T0) were compared on the last block of trials by means of Duncan's multiple-range test at a significance level of .05. In Fig. 1 the points contained within an enclosed line differ significantly from all those points outside of the enclosure, but do not differ significantly among themselves. Again, the principal effect involved T7 which showed more learning than any other task. Although no other adjacent tasks differed significantly, the pattern of performance differences in the last block of trials indicated that terminal performance was superior when the disparity of stimulus weights was greater. T1 was an exception to the foregoing generalization as it ranked in the middle of the tasks. However, the result for T1 was predicted since T1 required no discrimination of stimulus weights.

The significant quadratic component of the orthogonal comparison between tasks supports the prediction that performance will be related to disparity of weights by a bow-shaped function. The significant quadratic component of the interaction further supports this prediction.

Criterialities.—The correlation (Pearson r) of a stimulus dimension with S's responses is referred to as a *criteriality*, i.e., S's subjective weighting of the stimulus. It serves as a kind of diagnosis of what S has learned about each stimulus separately. Criterialities were determined for each stimulus dimension over successive blocks of 50 trials for each S. Figure 2 presents the mean criterialities by stimuli, blocks of trials, and tasks. An analysis of variance of trends was performed on the transformed criterialities for each stimulus dimension separately. Criterialities differed significantly between tasks in each of the three analyses, $F\,(6,\ 63) = 20.0$, 9.5, and 43.0, $p < .001$ for Stimuli 1, 2, and 3, respectively. This result confirms what can be seen by inspection of Fig. 2, i.e., the criterialities learned by *S*s in one task differed from those learned by *S*s in other tasks, and the direction of these differences was a function of the tasks' stimulus weights. There was a significant effect of blocks of trials in all three analyses, $F\,(2,\ 126) = 18.9$, 11.0, and 42.3, $p < .001$ for Stimuli 1, 2, and 3, respectively, indicating the learning of criterialities. In general, criterialities moved closer to the task requirements from block to block. The Tasks × Blocks interaction was significant in all three analyses, $F\,(12,\ 126) = 3.9$, $p < .05$ for Stimulus 1; $F\,(12,\ 126) = 5.7$, $p < .01$ for Stimulus 2; and $F\,(12,\ 126) = 7.8$, $p < .001$ for Stimulus 3. The amount of change in criterialities differed between tasks. No highly regular relationship between criteriality change and task appeared although there was a tendency for the maximum change to occur in tasks of intermediate disparity of weights (T3, T4, T5, and T6). Undoubtedly, this relationship was obscured by the rapid learning in the first block of trials in T5, T6, and T7.

Stimulus dependency.—The multiple correlation (R) between the three stimulus dimensions and *S*'s responses is referred to as *S*'s stimulus dependency. It provides a measure of the extent to which *S*'s responses depend upon the stimuli in a reliable manner, but it does not necessarily reflect the accuracy of his responding. The R may be high even when *S* has failed to weight the stimuli accurately. The *R*s were obtained over successive blocks of 50 trials for each *S*. An analysis of variance of trends of the transformed *R*s was conducted. Both tasks and blocks differed significantly, $F\,(6,\ 63) = 11.5$, $p < .001$, and $F\,(2,\ 126) = 42.4$, $p < .001$, respectively, while the Tasks × Blocks interaction was marginally significant, $F\,(12,\ 126) = 2.4$, $p < .05$. Tasks higher in disparity of weights generally exhibited higher *R*s, with the exception of a reversal between T2 and T1. A learning effect was shown in the increase in *R*s over blocks. The mean R for Tasks 1–7 on the third block of trials was $+.91$. The Tasks × Blocks interaction appears to be due to a relatively sharper increase in R for T7 and a drop in R in the third block for Tasks 3–6. The results of Duncan's multiple-range test (at the .05 significance level) between tasks on the last block of trials showed that T7 differed from all other tasks, T6 and T5 did not differ, T5 and T4 did not differ, and T4–T1 did not differ.

Even in T0, where stimuli were related randomly to criterion values, *S*s were responding to stimuli in a relatively nonrandom manner. The mean R for T0 on the third block of trials was $+.40$. Although T0 was not included in the preceding presentation of criteriality results, it should be noted that *S*s of T0 were highly heterogeneous in their criterialities. In the main, T0 criterialities were positive.

Error scores.—At first thought *S*'s error score, taken as the absolute difference between his response and the criterion, might appear to be the best single measure of interval concept learning. However, the error score is sensitive to at least three sources of variation, making its interpretation ambiguous: (*a*) the validity of *S*'s criterialities, (*b*) the variance

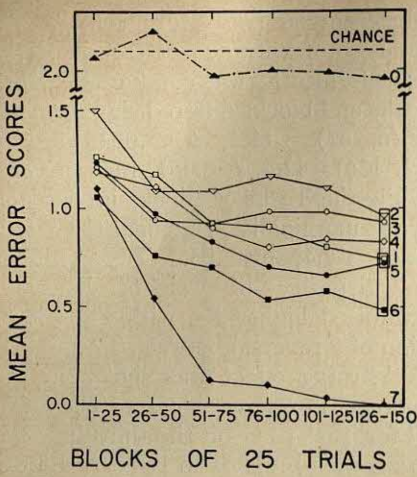


FIG. 3. Mean error scores of all tasks.

of the distribution of *S*'s responses (Cronbach & Gleser, 1957), and (*c*) the mean of the distribution of *S*'s responses. Nevertheless, the error score is a useful measure of *S*'s response accuracy which results from the combined effect of these three sources of variation. Six error scores, the mean error for each successive block of 25 trials, were obtained for each *S*. Figure 3 presents the mean error scores by tasks over the six blocks of trials. The lines enclosing points on the last block of trials represent the results of Duncan's multiple-range test at the .05 significance level. A one-way analysis of variance of trends was performed on the error scores. The error score results essentially parallel those for response validity. Tasks differed significantly, F (6, 63) $= 7.3$, $p < .001$; blocks differed significantly, F (5, 315) $= 41.8$, $p < .001$; and the interaction between blocks and tasks was significant, F (30, 315) $= 2.2$, $p < .01$. Mean error scores by tasks ranked from lowest to highest were: T7, T6, T5, T4, T1, T3, and T2. The significant interaction was due to the precipitous drop in error scores in T7. Task T7 differed significantly ($p < .01$) from all other tasks while none of the remaining tasks differed significantly ($p > .10$) from each other in rate of reduction of error scores.

The linear and quadratic orthogonal comparisons between tasks were significant, F (1, 63) $= 19.3$, $p < .001$, and F (1, 63) $= 5.8$, $p < .001$, respectively. Thus, the relationship between error scores and tasks was not monotonic, due to the lower error scores of T1 as compared with T3 and T2. Linear and quadratic comparisons between blocks were highly significant, F (1, 315) $= 170.3$, $p < .001$, and F (1, 315) $= 40.0$, $p < .001$, respectively, while the cubic comparison was marginally significant, F (1, 315) $= 6.2$, $p < .05$. The learning curve for all tasks combined was negatively accelerated and approaching an asymptote. Only the linear component of the Blocks × Tasks interaction was significant, F (6, 315) $= 7.4$, $p < .001$. This was another reflection of the large differences between T7 and all the remaining tasks.

Task T0 is included in Fig. 3 for the sake of comparing performance under random stimulus-criterion conditions with performance in Tasks 1–7. The horizontal line across the top of Fig. 3 represents a chance error score. The chance error score was calculated on the assumption of random responses with the mean and variance of the distribution of responses equal to that of the criterion distribution. The *S*s of T0 performed better than chance. Presumably, this was due to the significant reduction in the variance of their responses from 1.95 on the first block of 50 trials to 1.58 on the last block of 50 trials, F (2, 18) $= 4.1$, $p < .05$.

Interrelations of measures.—Error score was compared with response validity by means of a rank-order correlation (rho) for Tasks 1–7 separately over the last block of 50 trials. The median rho was +.69. Although the error score is sensitive to variations in the parameters of the response distribution, it showed close correspondence to the response validity measure.

The median rho (Tasks 1–7) between *R* and response validity was +.87 on the last block of 50 trials.

The rho between standard deviation of responses and error score (corrected for a departure of the mean response from the mean criterion) for those *S*s with a response validity less than +.71 was +.59 on the last block of 50 trials. The same rho for T0 was +.39.

Subjective probability estimates of criterion values.—At the end of the experimental session *S*s estimated the proportion of times each of the nine criterion values appeared. The median difference between the mean estimated proportion and the true proportion was .02 with a range of .00 to .05 for the nine criterion values. There is little question that one of the things *S*s learned was the distribution of criterion values.

DISCUSSION

The primary purpose of the present study was to demonstrate the feasibility of subjecting interval concept learning to experimental investigation and quantitative analysis. The results leave no question that the several partial contingencies between stimuli and predicted criterion events can be learned under these conditions. The extent of learning in this study was greater than that reported by Smedslund (1955) or by Summers (1962). This may have been due to a difference in *S*s or procedure or both. Summers' *S*s were ninth-grade children, and both Summers and Smedslund employed stimuli possibly more subject to error in identifying the values displayed.

The traditional concept experiment usually carries *S* to a criterion of perfect learning when the task makes this possible. The data in this study, as in the other studies of interval concept learning, indicate that *S*s' performances were approaching an asymptote definitely short of perfect or complete learning. The lack of a one-to-one correspondence between a stimulus value and a criterion value and the arithmetic process required in the weighting and combining of stimuli probably account, in part, for lack of perfect learning. Also, since criterion values were rounded to the nearest integer, one-unit errors were difficult for *S* to avoid even if his criterialities were accurate. This is analogous to the probabilistic feedback employed by Azuma (1960) and by Smedslund (1955) in interval concept studies and by Pishkin (1960) in a nominal concept study. Unlike the nominal concept study of R. E. Goodnow (reported in Bruner, Goodnow, & Austin, 1956, pp. 195–202) in which two partially relevant cues were used, it was possible for *S*s in Tasks 1–7 of this study to achieve errorless performance. This corresponds to the distinction made by Summers (1962) between an event frequency situation and a correlational situation. Whether or not there is a fundamental difference in the learning process in these two types of situations remains to be determined.

The structure of the stimulus, response, and criterion displays probably accounts for the fact that most *S*s were performing better than chance even in the first block of trials. The stimulus, response, and criterion dimensions formed parallel, linear scales with increments in scale value from left to right in conformance with most of the measurement scales in use in this culture. If *S* expected scale increments from left to right, this would have given him an initial advantage. Also, since all stimulus weights were positive (if not zero), *S* would have exceeded chance if he predicted the criterion value to move in the same direction as stimulus values. It seems likely this was the initial expectation of most *S*s. The great majority of T0 *S*s exhibited positive criterialities.

The effects of amount of disparity of stimulus weights agree well with classical studies of stimulus discrimination along a single physical continuum. Larger differences in stimulus weights were more readily discriminated as evidenced by more valid responding. Although the comparison of tasks adjacent in degree of disparity of weights did not satisfy conventional statistical criteria, the pattern of results supported the prediction that terminal response validity would be related to disparity of weights by a bow-shaped function. The data indicated that a unit increment in weight disparity had a greater effect upon learning the further removed from zero disparity the increment occurred. Investigation of a number of points of weight disparity between that which occurred in T1 and T3 will be required before a precise specification can be made of the point of inflection in the functional relationship between disparity of weights and performance.

The implication seems clear that accurate prediction will be difficult to attain when based upon multiple stimuli which differ only moderately in validity. In addition to the problem of discriminating moderate weight differences, several stimuli must be attended to even if the necessary weight discriminations are made. Pollack (1953) showed that accuracy of stimulus information transmission suffered as the number of stimuli simultaneously displayed increased. An increase in the number of stimuli requiring attention is likely to have the same effect. Also, the arithmetic task is more difficult when several valid stimuli as opposed to only two or one must be combined in deciding upon a response.

The data show that probability learning is involved in the interval concept task in addition to the learning of relationships between stimuli and criterion. The subjective probability estimates of criterion values showed that *S*s had learned the probability distribution of criterion values. The reduction in error scores by T0 *S*s only could have been due to their response variance reduction, which interpretation the rho between error score and response variance further supports. The same relationship was found among *S*s in Tasks 1–7 with low response validities. There can be no doubt that learning the statistical properties of a set of predicted events can improve predictive accuracy when the predictive scheme itself, i.e., set of criterialities, is not highly valid. However, in the case of T0, no *S* limited the variance of his responses to the optimum, i.e., zero. The behavior of T0 *S*s approximated probability matching analogous to that frequently observed in two-choice probability learning studies (e.g., Estes & Straughan, 1954). Extrapolating from the Estes (1957) model, equality of response and criterion variance would be a case of probability matching when all stimuli have zero validity. However, there was no sign that the response variance of T0 *S*s had ceased to decrease at the end of training. Whether the tendency of their responses to match criterion probabilities would have continued over an extended series of trials must remain problematical for the time being.

REFERENCES

Adams, J. Concepts as operators. *Psychol. Rev.*, 1953, **60**, 241–251.

Azuma, H. Comparison of a correlational with a probabilistic approach to concept learning. Unpublished doctoral dissertation, University of Illinois, 1960.

Bourne, L. E., Jr. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

Cronbach, L. S., & Gleser, G. C. *Psychological tests and personnel decisions.* Urbana: Univer. Illinois Press, 1957.

Estes, W. K. Theory of learning with constant, variable, or contingent probabilities of reinforcement. *Psychometrika*, 1957, **22**, 113–132.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Hoffman, P. J. Generating variables with arbitrary properties. *Psychometrika*, 1959, **24**, 265–267.

Hull, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**(1, Whole No. 123).

Pishkin, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *J. exp. Psychol.*, 1960, **59**, 371–378.

Pollack, I. The information of elementary auditory displays. II. *J. Acoust. Soc. Amer.*, 1953, **25**, 765–769.

Reed, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. exp. Psychol.*, 1946, **36**, 71–87.

Smedslund, J. *Multiple probability learning.* Oslo: Akademick Forlog, 1955.

Stevens, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Ch. 1.

Summers, S. A. The learning of responses to multiple weighted cues. *J. exp. Psychol.*, 1962, **64**, 29–34.

Uhl, C. N. An analysis of dimensional judgment concepts. Unpublished doctoral dissertation, University of Oregon, 1960.

(Received September 29, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 274–280

LEARNING OF RESPONSES TO STIMULUS CLASSES AND TO SPECIFIC STIMULI[1]

BURTON H. COHEN AND PETER A. HUT

Lafayette College

To test Miller's chunk hypothesis, the performance of Ss who learned a common number response to each of 8 categories of 3 words was compared with that of Ss who learned control lists of 8 word-number pairs. Performance of Ss who learned number responses to Exhaustive word categories, i.e., categories whose 3 words represent all (or nearly all) the words belonging to the category, did not differ significantly from that of control Ss. Performance of Ss who learned number responses to Nonexhaustive word categories, i.e., categories whose 3 words represent a small sample of the words belonging to the category, differed significantly ($p < .001$) from both the control Ss and Ss presented with the Exhaustive categories. The results were interpreted as providing support for Miller's views regarding verbal recoding.

Using a paired-associate (PA) learning task, Lloyd (1960) found support for the hypothesis that there would be no difference in the retention of responses to classes of stimuli compared to the retention of responses to specific stimuli—provided the number of responses that were learned was the same in both cases. One group of *S*s, the *class* group, received a list of 24 S-R pairs whose stimuli could be grouped into categories of three words each. Another group of *S*s, the *specific* group, learned a list of 8, specific, pairs where the same 8 pairs were repeated in three random orders to form a list of 24 pairs. No significant difference was found between the two groups on a test of retention.

With respect to original learning, however, the class group required significantly more trials to criterion than the specific group. This result was not unexpected since the class group had to handle three times the number of stimuli as the specific group. However, according to Miller's (1956) discussion of verbal recoding, a group of related items may be recoded by assigning a common label or name to all the items. Thus, instead of handling each item on an individual basis, the recoded name or label is processed and handled in memory. Viewed in this manner, the 24 pairs of Lloyd's class lists could theoretically function as an 8-pair list—provided the recoding operation was simple and efficient. It was evident that Lloyd's categories did not meet these criteria. If classes of words which can be simply and efficiently recoded are substituted for Lloyd's classes, it should be possible to demonstrate that a class group requires no more learning trials than a specific group.

Cohen (1963) has drawn a distinction between two kinds of word categories. On the one hand, the categories that Lloyd used are called nonexhaustive (NE) since the selected words represent only a small fraction of the words which nominally belong to the category. On the other hand, there are instances of categories where three words completely or almost completely exhaust all the words that belong under a category label, e.g., MASCULINE, FEMININE, and NEUTER. Such categories are called exhaustive (E). When E and NE categories were compared in a free recall situation, Cohen (1963) found that E categories were handled in a more unitary fashion than NE categories.

---

[1] This research was supported, in part, by Research Grant M-4430(A) from the United States Public Health Service, the National Institutes of Health and by the National Science Foundation's summer program of research participation for college teachers held at the Pennsylvania State University.

The present study compared the learning of an E and an NE class group with their appropriate, specific, control groups. If *S*s can handle each E category as a single unit or chunk, it should take no longer to learn 24 pairs comprising eight chunks or categories of three words each than it takes to learn a list of 8, specific, S-R pairs.

EXPERIMENT I

Procedure

The design of this study was essentially the same as the original learning task employed by Lloyd. Four PA lists of 24 word-number pairs were used. For all lists, the response numbers were the digits 1 to 8. Two of the lists were class lists in that the stimuli of the 24 pairs comprised eight categories of three words each. For a given *S*, all three words of a given category were paired with the same digit. Thus, for the class lists, *S*s were presented with 24 different stimuli but the same eight digits were repeated three times. The other two lists were specific lists in that they comprised the same eight pairs repeated three times. By defining a trial as the presentation of 24 pairs, the number of response units to be learned was equal for both types of list.

The stimuli of the PA lists comprised two sets of words. One set was the eight, three-word, NE categories used by Lloyd (1960).[2] The other set contained eight, three-word, E categories selected from the stimulus lists the senior author had used in previous free recall studies. Because of the limited number of three-word E categories from which selection could be made, no specific attempt was made to match the categories or words of the two word sets. (Lloyd found no relation between the mean number of trials to criterion and Thorndike-Lorge frequency for his lists.) The eight E and eight NE categories are shown in Table 1. All 24 words of the NE categories were used to form a class nonexhaustive (CNE) list. Specific nonexhaustive (SNE) lists were formed by selecting one word from each of the eight NE categories. In a similar fashion, class exhaustive (CE) and specific exhaustive (SE) lists were formed.

TABLE 1

STIMULUS WORDS

Exhaustive		Nonexhaustive	
MASCULINE FEMININE NEUTER	BLONDE BRUNETTE REDHEAD	SQUARE TRIANGLE ELLIPSE	PIGEON LARK ORIOLE
VOLT AMPERE OHM	SOLDIER SAILOR MARINE	COAT JACKET SMOCK	PRIEST MONK DEACON
CIGARETTE CIGAR PIPE	KNIFE FORK SPOON	PINE ELM HICKORY	TABLE COUCH FOOTSTOOL
LENGTH WIDTH HEIGHT	CATHOLIC PROTESTANT JEW	AIRPLANE WAGON SAILBOAT	TROUT SALMON BASS

For the CE and CNE lists, eight different randomizations of the 24 word-number pairs were drawn such that: (*a*) each category was systematically paired with every number; (*b*) within each subtrial of eight pairs, every category was represented once; and (*c*) six different random orderings of the pairs of each subtrial were drawn to produce two different presentation orders for the 24 pairs. For the SE and SNE lists, 24 different randomizations of the 24 pairs were drawn such that: (*a*) each of three subsets of eight lists contained one word from each of the eight categories; (*b*) within each subset, each word was systematically paired with every number; and (*c*) six different random orderings of the pairs of each of the 24 randomizations were drawn to produce two different presentation orders of 24 pairs. In no case was the last pair of a subtrial of any list identical with the first pair of the next subtrial.

The pairs were typed (words in capital letters) on adding machine tape and presented to *S*s by means of a memory drum similar to that described by Adamson (1955). Replicating the procedure used by Lloyd, the pairs were presented at a 2:2 sec. rate with an intertrial interval of 24 sec.

The *S*s were 96 male, undergraduate,

---

[2] The writer is indebted to Kenneth E. Lloyd for providing a copy of his doctoral dissertation.

Lafayette College students who were randomly assigned to provide a total of 24 *S*s per condition. For the CE and CNE conditions, 3 *S*s were assigned to each of the eight different list randomizations. For the SE and SNE conditions, 8 *S*s were assigned to each of the three subsets of eight, specific, S-R pairs.

The *S*s were given standard PA learning instructions. Since some *S*s had served previously in PA learning studies, several trials on a short list of nonsense syllable-letter pairs were given to *S*s who were naive concerning PA learning. The *S*s in the class conditions were not told that three different words would be paired with the same response digit. All *S*s were informed that the learning criterion was one perfect trial. At this point, one major change from Lloyd's procedure was introduced. The *S*s for all of Lloyd's original learning conditions were given approximately 1 min. to study a sheet of paper which contained all the stimulus words grouped into their appropriate categories. The *S*s of the present experiment were not shown the stimulus words.

Results

The means and *SD*s for the number of trials to criterion for each of the four conditions are shown in Table 2. The results of the 2 × 2 analysis of variance were as follows: for List Type (specific vs. class), $F\ (1,\ 92) = 19.02$, $p < .001$; for Category Type (NE vs. E), $F\ (1,\ 92) = 7.77$, $p < .01$; and for List Type × Category Type interaction, $F\ (1,\ 92) < 1.0$. Since the experimental hypothesis specified that the CE and SE means should be similar and since there was no reason to expect a significant difference between the SE and SNE means, support for the hypothesis would be derived from a significant interaction term. The fact that the interaction term was not significant was clearly contrary to the hypothesis under investigation.

TABLE 2

MEANS AND *SD*s OF TRIALS TO CRITERION FOR EXP. I

List Type	Category Type	*M*	*SD*
Specific	Exhaustive	4.00	1.87
	Nonexhaustive	5.79	3.30
Class	Exhaustive	6.79	2.57
	Nonexhaustive	8.54	4.02

In general, the results of Exp. I were quite unstable. Both the means and *SD*s of the three subsets of the SNE list were significantly different ($p < .05$ in both cases). Similarly, Bartlett's test, applied to the variances of the four main lists, showed a significant departure from homogeneity ($p < .001$). In view of the negative and ambiguous nature of these results, it was decided to treat Exp. I as a pilot study and to replicate the study with several changes.

EXPERIMENT II

Procedure

With the exception of the following changes, the procedure for Exp. II was identical to that of Exp. I: (*a*) Only *S*s naive with respect to PA learning were tested; (*b*) A Lafayette memory drum, Model 303B, was substituted for the memory drum used in Exp. I; (*c*) The three words of all 16 E and NE categories, grouped into their appropriate categories, were shown for 1 min. to all *S*s. Eight different random orders of the 16 categories were used; (*d*) A modified set of instructions for PA learning was used.

Briefly, the procedure for Exp. II was as follows. Paired-associate learning was verbally described to *S*s. Then, several practice trials on a list of five nonsense syllable-letter pairs were administered. The *S*s were informed that the experimental list comprised words as stimuli and that the responses were the digits 1 to 8. They were told to expect a long list of S-R pairs and were informed that a fairly long interval would separate the trials. Heavy emphasis was placed on instructions to: (*a*) always call the response digits out loud; (*b*) always anticipate if the correct response digit was known; and (*c*) either guess or remain silent to stimuli for which the response had not been learned. An attempt was made to implement first-trial responding by instructing *S*s to call out responses "as soon as you reach the stage where you think you know the response which is paired with a given stimulus." Finally, *S*s were shown the list of 16, three-word categories, were told to study the list for 1 min., and were told that some of the words on the study sheet would be identical with the stimuli of the list they were about to learn. Immediately after the 1-min. study period, the PA learning began.

The *S*s were 96 male, undergraduate students attending the summer term at the Pennsylvania State University. Assignment of *S*s to the four conditions was identical to that employed in Exp. I.

In addition to the PA learning task, word associations to the 24 E and 24 NE words were collected. Cohen (1962, 1963) has shown that an index of interitem associative strength (IAS), derived from word association norms, effectively discriminated between E and NE categories and, in a free recall task, the IAS index correlated significantly with within-category word recall. In view of these findings, it appeared desirable to assess the ability of the IAS index to predict learning differences among the E and NE categories of the present study. The technique used in collecting the norms has been described in detail by Cohen (1963). For both the E and NE words, eight different orders of the 24 words were used and 7 *S*s were randomly assigned to each order. One hundred and twelve, male, undergraduate students attending the summer term at the Pennsylvania State University served as *S*s—a total of 56 *S*s on each class list. None of the *S*s had participated in the PA learning task.

Results

The means and *SD*s of the number of trials to criterion for the four conditions are shown in Table 3. (There were no significant differences between the means or the variances for the three subsets of the SE and SNE lists.) Bartlett's test revealed no significant departure from homogeniety of variance for the four conditions ($.80 < p < .90$). The results of the 2 × 2 analysis of variance were as follows: for List Type, $F(1, 92) = 19.51$, $p < .001$; for Category Type, $F(1, 92) = 6.01$, $p < .025$; and for List Type × Category Type interaction, $F(1, 92) = 7.59$, $p < .01$. In view of the significant interaction term, attention was directed to the evaluation of the simple effects. Application of a test of multiple contrasts suggested by Dixon and Massey (1957, pp. 169–171) showed that there were no significant differences between the means of the SE, SNE, and CE conditions while all differences between these conditions and the CNE condition were significant at better than the .01 level. These results provided strong support for the experimental hypothesis.

TABLE 3

MEANS AND *SD*s OF TRIALS TO CRITERION FOR EXP. II

List Type	Category Type	*M*	*SD*
Specific	Exhaustive	5.08	2.41
	Nonexhaustive	4.92	2.76
Class	Exhaustive	6.00	2.80
	Nonexhaustive	8.88	2.82

While the trials to criterion measure may be an adequate measure of learning for the class lists, it is quite a gross measure for the specific lists since the same eight pairs were repeated three times within a trial. Thus, the nonsignificant difference between the means of the CE and SE conditions may be a spurious result of the grossness of the trials to criterion measure. To evaluate this possibility, two finer measures of learning performance were derived from the learning protocols. The first measure, designated 8F, represents the first subtrial of eight pairs on which all the responses were correctly anticipated. The second measure, designated 8L, represents the subtrial immediately following that subtrial on which the last error was made. Table 4 presents the means and *SD*s of these two measures for the four conditions of Exp. II. The analyses of variance tests applied to these two learning measures showed the same pattern of results as the trials to criterion measure. Likewise, application of the Dixon and Massey multiple-contrasts test to evaluate the simple effects showed the same pattern as before. In all respects, the 8F and 8L measures confirmed the results obtained from the trials to criterion measure.

Though nonsignificant in every case, all three learning measures showed a larger mean for the CE than for the SE condition. However, an adjustment may be applied to the CE means which reduces the absolute magnitude of these differences. In a standard PA learning situation, no anticipations, correct or otherwise, occur until *S*s have seen the entire list once. However, the instructions used in Exp. II attempted to force *S*s to anticipate on Trial 1. This was done so that *S*s of the SE and SNE conditions would receive credit for correct responses on the second subtrial of Trial 1 since this subtrial would normally be Trial 2 of a standard PA list. In contrast, since all 24 S-R pairs had to be presented before *S*s saw the entire class list, these *S*s would not be expected to anticipate until Trial 2. As a result, class *S*s were penalized on measures of learning.

The magnitude of the penality was assessed by determining the mean pair number on which the first anticipation, correct or otherwise, was made. (A value of 25 was arbitrarily assigned if no anticipations were made on Trial 1.) The mean pair numbers were 11.58, 11.08, 18.79, and 20.63 for the SE, SNE, CE, and CNE conditions, respectively. On the average, *S*s of the CE and CNE conditions started anticipating approximately eight pairs after *S*s of the SE and SNE conditions, i.e., about one subtrial later. If each mean for the CE condition in Table 4 is reduced by one, the absolute magnitude of the difference between the CE and SE conditions is reduced by approximately 50% for the 8F measure and 33% for the 8L measure.

Figure 1 shows the mean number of correct responses for each of the first 15 subtrials. Because of highly similar curves for the SE and SNE conditions, the sets of data were pooled. The curves presented in Fig. 1 give graphic evidence of the discrepancy between the CNE and the other three conditions. Also, in

TABLE 4

MEANS AND *SD*s OF 8F AND 8L LEARNING MEASURES FOR EXP. II

List Type	Category Type	Learning Measure			
		8F		8L	
		M	*SD*	*M*	*SD*
Specific	Exhaustive	8.21	4.57	12.33	6.97
	Nonexhaustive	8.04	4.28	12.21	8.09
Class	Exhaustive	10.04	4.65	15.21	8.04
	Nonexhaustive	17.96	7.54	23.83	8.15

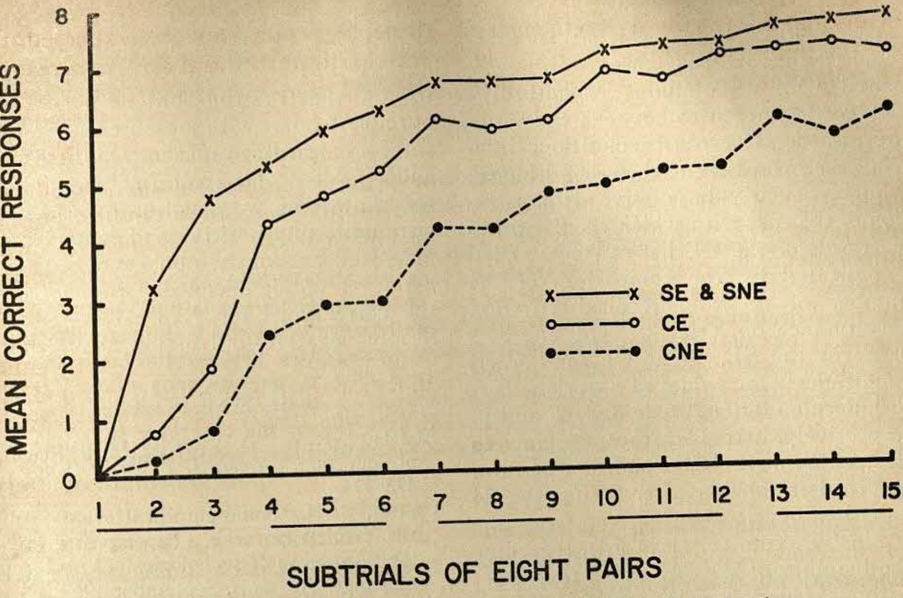


FIG. 1. Mean number of correct responses per subtrial of eight pairs.

accordance with the penalty discussed above, if one visualizes that each point for the CE condition is displaced one subtrial to the left, the correspondence between the curves for the CE and the combined SE-SNE conditions is considerably improved. It is interesting to note that the CE and CNE curves show an initial period of positive acceleration while the SE-SNE curve is negatively accelerated throughout.

The relative ease with which *S*s handled the CE compared to the CNE list was directly related to the IAS index. As described by Deese (1959), the IAS index measures the degree to which the three words of each E and NE category elicit each other as word associates. The mean IAS indices for the eight E and eight NE word categories were 29.24 and 6.47, respectively. In only one instance was the IAS index for an NE category as high as the lowest IAS index for an E category. When, however, a variety of measures of within-category learning efficiency for the E and NE categories were correlated with the IAS index, no significant relations were found.

DISCUSSION

The results of Exp. II provided strong evidence for the hypothesis that the 24 words of the CE condition could be handled as eight packages or chunks of information. The consistent but nonsignificant differences between the SE and CE conditions may be attributed to two factors. First, the class *S*s had to discover for themselves that three words were paired with the same response. Second, even though all *S*s were exposed to the stimulus words prior to PA learning, it was probably the case that the 1-min. study period did not provide sufficient time for the class *S*s to fully grasp and retain the groupings of the words. Thus, at the start of PA learning, the class *S*s (especially those of the CNE condition) still had to learn to discriminate among the stimulus words and to group them into the eight, appropriate categories. Since *S*s of the CE and CNE conditions would not be expected to differ in the length of time required to learn that three words were paired with a common response, the significant differences between the learning measures for these conditions indicates the operation of the second factor.

A particularly impressive demonstration of the ease with which the 24 stimulus words of the CE list can be recoded is indicated by the comparison of the number of *S*s who reached criterion performance on Trial 2. The numbers of *S*s in the SE, SNE, CE, and CNE conditions who achieved errorless performance on Trial 2 were 3, 4, 2, and 0, respectively. It is rather striking to observe *S*s of the CE condition perform in errorless fashion after seeing the 24 pairs just once.

Finally, a comment is in order regarding the inability of the IAS index to predict within-category learning efficiency. Both Deese (1959) and Cohen (1963) have shown that the IAS index is an efficient predictor of within-category word recall. It is necessary to point out that in both these experiments the results were obtained in a free recall situation where *S*s had no a priori knowledge of the stimulus words. In general, it would not be expected that the change from free recall to a learning task would effect the overall ability of the IAS index to predict learning differences among the word categories. However, it is possible that showing *S*s the words grouped into their appropriate categories prior to the start of learning has the effect of attenuating the value of the IAS index as an effective learning cue. It is interesting to note that the rank-order correlations of the IAS index with measures of within-category learning efficiency for the E and NE data of Exp. I (where *S*s were not shown the words prior to learning) were uniformly positive and of moderate magnitude, though none reached the .05 level of significance. In contrast, the correlations for Exp. II were uniformly negative and of low magnitude.

REFERENCES

ADAMSON, R. A. A drum for tachistoscopic presentation. *Amer. J. Psychol.*, 1955, **68**, 473–474.

COHEN, B. H. Some empirical evidence for the chunk hypothesis. Paper presented at the American Psychological Association meetings, St. Louis, September 1962.

COHEN, B. H. An investigation of recoding in free recall. *J. exp. Psychol.*, 1963, **65**, 368–376.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis.* New York: McGraw-Hill, 1957.

LLOYD, K. E. The retention of responses to stimulus classes and to specific stimuli. *J. exp. Psychol.*, 1960, **59**, 54–59.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

(Received October 1, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 281–285

BETA MOTION THRESHOLDS

FRANK J. SGRO[1]

Rutgers University

3 men and 2 women (normal Ss) were tested for beta motion thresholds with variations of interval and duration times from 5 to 100 msec. Ss were given preliminary training in beta motion perception for the purpose of avoiding naive judgments resulting from inability to distinguish true beta motion from other types of apparent motion. The apparatus consisted of the RCD4 Apparent Motion Unit, a specially built unit featuring 2 vertically mounted lights with various speeds of duration and interval time with discrete gradations as fine as 1 msec. The results show compliance with Korte's 4th law below 65-msec. values; interval time decreases as duration time increases. Above 65 msec. the values do not comply with Korte's law. As duration time is increased the function approaches an asymptote, indicating a failure of Korte's 4th law.

The efficacy of apparent motion and flicker perception in diagnosing clinical groups has been investigated by a number of authors with the purpose of assessing the severity of brain damage and neural malfunction by means of ability to perceive apparent motion (Saucer, 1942, 1958; Saucer & Deabler, 1956; Teuber & Bender, 1948; Werner & Thuma, 1942). In essence, the above authors indicate that apparent motion perception may be an invaluable aid as a diagnostic instrument in classifying different clinical groups on the basis of the ability to perceive apparent motion.

Although there is a large amount of material in the literature on apparent motion perception, no one has attempted to establish thresholds for normals to aid in distinguishing normal groups from clinical groups. This has left an unfortunate gap in the literature.

The present experiment was designed to determine the thresholds for beta motion perception for normal *S*s by means of a systematic investigation of variations of duration and interval time of the presentation of two vertical lights, using the method of limits. Beta motion is defined as apparent movement wherein the two lights seem to fuse into one light which appears to move (Graham, 1951, p. 899).

METHOD

Apparatus.—The Apparent Movement Unit RCD4 is a custom built unit manufactured by the Psychological Instruments and Research Laboratory of Montreal, Canada. The unit is designed to provide fast rise and decay rates of two stimulus lights vertically mounted on a flat surface. The timing unit allows for the variation of duration and/or interval time from 1 msec. to 5 sec. The timing circuit is a cathode-coupled monostable multivibrator with positive grid return. The duration of time on and/or interval between lights is continuously variable at three different settings. Each setting can be manipulated at any speed including discrete gradations of 1 msec. The RCD4 Unit is the first of its kind built.

The viewing panel was 45 × 35 cm. and was covered with black paper. The two

---

[1] This study was conducted while the author was an intern at the Eastern State Hospital, Williamsburg, Virginia. The author gratefully acknowledges Eastern State Hospital for the use of their facilities, and J. L. Chambers for his criticism of the original manuscript. Further acknowledgment is extended to the staff of the Psychology Department of the College of William and Mary for their pertinent suggestions regarding the experimental design.

TABLE 1

INTERVAL AND DURATION RANGES DERIVED FROM PRACTICE TRIALS

Duration Constant (msec.)	Interval Range (msec.)	Interval Constant (msec.)	Duration Range (msec.)
5	60–140	5	100–190
15	50–120	15	90–180
25	40–110	25	80–170
35	30–90	35	60–150
45	20–80	45	40–120
55	10–70	55	15–60
65	10–50	65	15–50
80	10–45	80	12–40
100	5–45	100	2–26

Note.—All interval and duration ranges were measured in 10-msec. gradations except for the 65- and 80-msec. intervals, where duration gradations were 4 msec., and for the 100-msec. interval where duration gradations were 2 msec.

stimulus lights were centered on the viewing panel. The lights were covered with red plastic lenses (provided by the Dialight Corporation, Brooklyn, New York) and emitted a wave length of from 650 to 710 mμ. The lights were vertically mounted, 10½ cm. apart from center to center, and the diameter of the lights with plastic caps was 1.27 cm. The lamps were ¼-w. NE 51 Neon Glow lamps with an ionization and deionization time of less than ½ μsec. The walls and ceiling of the room were painted a flat black and the room was completely darkened during the testing period. The viewing panel was 358 cm. from *S*, at eye level, subtending a visual angle of 1.7°. The luminance value of one light was 135 mL. at the source.

Subjects.—The *S*s were five trained observers from the staff of the psychology department of the Eastern State Hospital, Williamsburg, Virginia. There were 3 men and 2 women, all with normal or corrected vision. Training consisted of a complete series of preliminary testing with full ranges of duration time and interval time settings to familiarize *S*s with the perception of beta motion. The time values for presentation in the experiment proper were also derived from the preliminary threshold ratings.

Procedure.—When brought to the testing room, *S* sat in complete darkness for 2½ min. to complete the dark adaptation process. The *S*s were tested for thresholds of duration and interval time at the values of: 5, 15, 25, 35, 45, 55, 65, 80, and 100 msec.; i.e., duration threshold as a function of these interval times, and interval threshold as a function of these duration times.

The *S* was given one testing period per day for 15 days, intermittently over a period of 5 wk. The presentation of stimuli was randomly selected for each day, and the starting point for each day was also randomized. An ascending and descending stimulus presentation was used in an alternate ADDA pattern to compensate for the constant errors of expectation and habituation. Every value tested was begun at least two steps above or below *S*'s threshold (as determined by practice sessions in preliminary testing), with the following ranges used as presented in Table 1.

Values 5, 15, 25, 35, 45, and 55 were tested once per day for 15 days. Values 65, 80, and 100 were tested on only 3 days: Days 5, 10, and 15. The last three values were used only to get some indication of what occurs at the high value level. All presentations were discrete, at the gradations noted. Each value was presented for an exposure time of 3 sec. with a 5-sec. interval between presentations. A criterion of three consecutive affirmative or three consecutive negative responses was used; i.e., *S* had to report "yes" for three consecutive settings for seeing beta motion and "no" for three consecutive settings for not seeing beta motion.

The *S* was given the following instructions on every testing day while completing dark adaptation:

> In a few minutes I am going to show you two lights. I'd like you to look at these lights and tell me when the two lights seem to fuse into one single light moving up and down. I will ask you during each exposure if you see the movement of one light: answer yes or no. Do not blink your eyes while the lights are on; they will be on for only a short duration.

No instructions were given regarding fixation at any specific area.

The *S* was told not to blink his eyes as a precaution to prevent him from losing the motion perception, as preliminary investigation during practice sessions demonstrated that the movement perception may be lost if the eyes are opened and closed. The 5-sec. interval between presentations was also used to prevent unnecessary eye fatigue, which was considered to be an important factor as suggested by the studies of Brenner (1953), Crutchfield and Edwards (1949), and DeSilva (1926), reporting that continuous stimulation breaks down the apparent motion perception as a result of visual fatigue.

As each exposure was presented, *E* scored

S's answer on the score sheet at the appropriate millisecond level. The duration of each testing period was approximately 25 min., except for Days 5, 10, and 15, when the additional three values were tested. On those days, the testing period ran for another 10 min.

RESULTS AND DISCUSSION

The stimulus thresholds for duration and interval time were determined by averaging the ascending and descending thresholds for each duration and interval setting for each *S* for the 15 days. A group mean was then computed for overall thresholds for each duration and interval time. The *SD*s were computed from the grand means of the ascending and descending thresholds. A related *t* test was used to determine the significance of difference between the largest spread of ascending and descending values for duration and for interval. For duration constant at 55 msec. (the largest difference in ascending and descending) a *t* of 1.09 was found. For the interval constant at 15 msec. (the largest difference for interval) a *t* of .60 was found. With $df = 4$, neither *t* was significant, suggesting that the ascending and the descending means are not different. Table 2 gives a summary of the group means and *SD*s.

Except for interval constant at 55 msec. and 65 msec., there is no reversal of diminution of the means as duration is increased, demonstrating compliance with Korte's fourth law (duration decreases as interval increases, intensity and distance held constant: Graham, 1951, p. 900). The reversal of the means at 55 and 65 msec. could be due to the interval values being too close together to afford a clear difference, indicating that the experimental design should have provided for a much wider spacing of intervals and an extension of the range above the 55-msec. level.

Beyond the 55-msec. level, the curve begins to decelerate, suggesting a limitation to the range to which Korte's fourth law applies. Therefore it would be necessary to examine higher values to determine the limitations of Korte's fourth law.

Graham (1951) offers another possible explanation for the phenomenon: "Korte and Neuhaus have shown that distance between stimuli, a basic variable, must be increased as the pause between stimuli increases [p. 901]." This would suggest that any value higher than 55 msec. must be tested with an angular separation greater than 1.7° in order to afford optimal movement perception.

Figures 1 and 2 give a graphic representation of the group means of

TABLE 2

RANGE, MEANS, AND *SD*s OF DURATION AND INTERVAL VALUES

Duration Constant (msec.)	Range	*M*	*SD*	Interval Constant (msec.)	Range	*M*	*SD*
5	85–125	102	4.24	5	115–175	137	5.65
15	75–102	86	3.32	15	95–145	119	4.60
25	55–85	71	3.00	25	85–135	107	5.91
35	45–75	57	2.45	35	65–115	89	4.56
45	35–55	48	1.42	45	45–95	75	6.48
55	35–55	42	1.42	55	22–42	30	.995
65	32–47	39	1.73	65	29–42	34	2.23
80	30–42	35	.995	80	14–34	25	4.70
100	22–42	31	2.83	100	9–22	15	.995

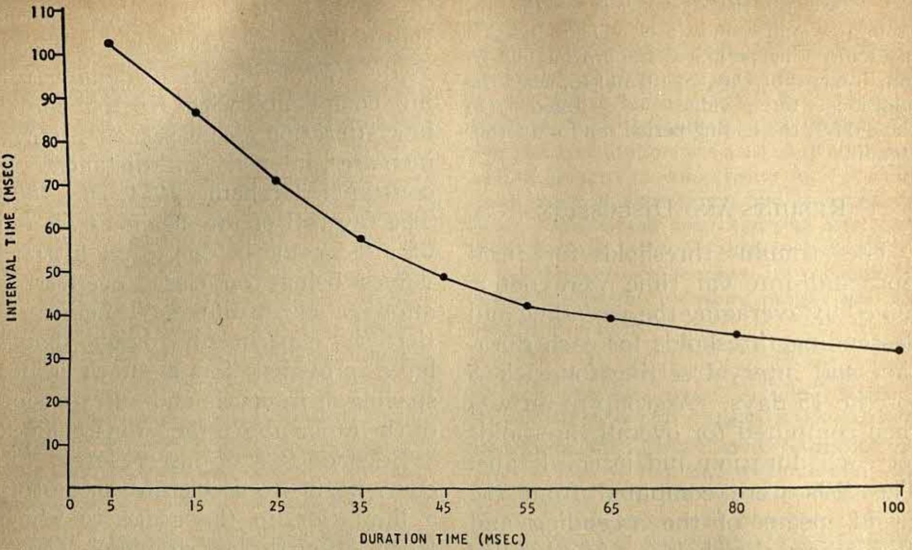


FIG. 1. Interval threshold as a function of duration time (group mean).

Table 2. Figure 1 shows beta motion interval thresholds as a function of duration time. Figure 2 shows beta motion duration thresholds as a function of interval time. Figure 1 suggests that beta motion is much more easily perceived with duration constant and interval varied, which is

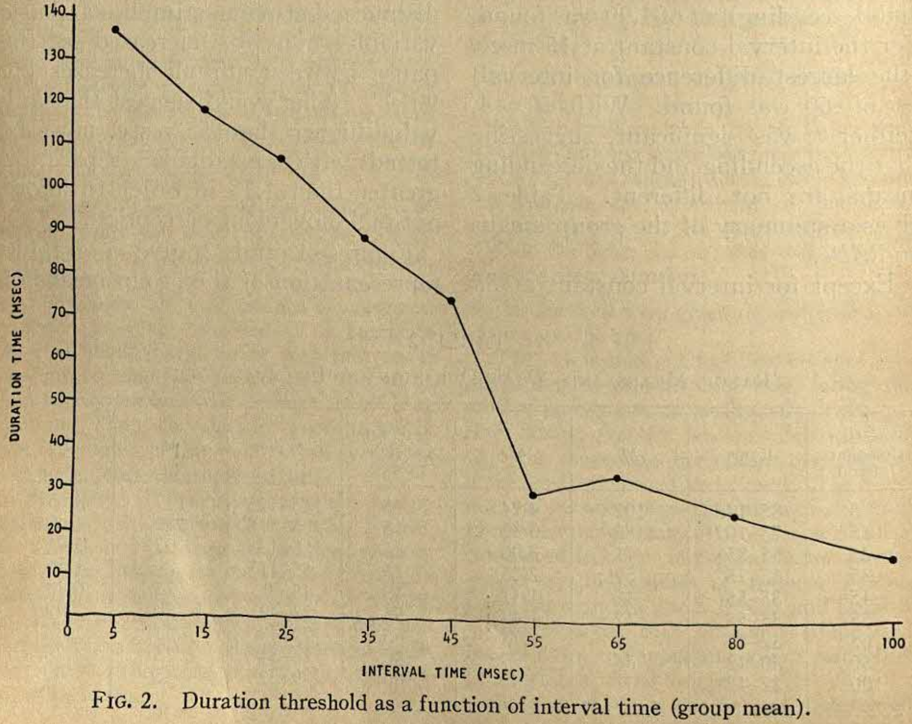


FIG. 2. Duration threshold as a function of interval time (group mean).

consistent with *S*s' remarks noting that there was "less strain" with duration constant as opposed to squinting and eye fatigue with interval constant. The function tends to become asymptotic if distance between stimuli, and distance between stimuli and *S* are held constant, and the stability of the function would most likely continue at higher values.

Further inspection of Fig. 1 also shows that Wertheimer's interval figure of 60 msec. for the occurrence of optimal movement (Graham, 1951, p. 900) is questionable unless specific angular separation and duration time are provided for in order to perceive optimal movement. For example, Fig. 1 shows that with an interval time of 60 msec., a duration setting of approximately 32 msec. is required (and this is with an angular separation of 1.7°), for the perception of optimal movement.

The extent of practice effect may also be an important factor in beta motion perception. If the thresholds are a function of learning, then proper interpretation of the data must be made to assess the value of the experimental design in view of the prime purpose. By inspection, the means for each value of duration time and interval time indicated that there was no practice effect of any significance over the 15-day period.

REFERENCES

BRENNER, M. W. Continuous stimulation and apparent motion. *Amer. J. Psychol.*, 1953, **66**, 494–495.

CRUTCHFIELD, R. S., & EDWARDS, W. The effect of a fixated figure on autokinetic movement. *J. exp. Psychol.*, 1949, **39**, 561–568.

DESILVA, H. R. An experimental investigation of the determinents of apparent visual movement. *Amer. J. Psychol.* 1926, **37**, 469–501.

GRAHAM, C. H. Visual perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 868–916.

SAUCER, R. T. A further study of the perception of apparent motion by schizophrenics. *J. consult. Psychol.*, 1942, **55**, 256–258.

SAUCER, R. T. Perception of the shortest noticeable dark time by schizophrenics. *Science*, 1958, **127**, 698–699.

SAUCER, R. T., & DEABLER, H. L. Perception of apparent motion in organics and schizophrenics. *J. consult. Psychol.*, 1956, **20**, 385–389.

TEUBER, H-L., & BENDER, M. B. Changes in visual perception of flicker, apparent motion, and real motion after cerebral trauma. *Amer. Psychologist*, 1948, **3**, 246–247.

WERNER, H., & THUMA, B. A deficiency in the perception of apparent motion in children with brain injury. *Amer. J. Psychol.*, 1942, **55**, 58–67.

(Received October 1, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 286–291

EFFECT OF NUMBER OF ACQUISITION TRIALS AND THE PRESENCE OR ABSENCE OF THE UCS ON EXTINCTION OF THE EYELID CR[1]

KENNETH W. SPENCE, EDWARD F. RUTLEDGE,
AND JOHN H. TALBOTT

State University of Iowa

Extinction of the eyelid CR in 100 Ss was studied as a function of number of acquisition trials (32 or 64) and the type of extinction procedure: (a) no UCS or (b) an extended CS-UCS interval of 2,500 msec. In agreement with earlier studies that had employed continuous reinforcement there was no significant difference in the rate of extinction as a function of extinction procedure. Extinction was significantly related to number of original conditioning trials. The results were interpreted in terms of a discrimination hypothesis that made use of the concept of set.

After first ascertaining that little or no eyelid conditioning occurred in human *S*s when an extended CS-UCS interval of 2,500 msec. is employed, McAllister (1953a, 1953b) subsequently demonstrated that CRs established at an optimal interval (500 msec.) extinguished when the CS-UCS was increased to 2,500 msec. Comparing extinction under this procedure with the traditional one of omitting the UCS, McAllister found that the decrement in response strength fell significantly slower under the new procedure. He interpreted this greater resistance to extinction as being primarily due to the fact that the level of drive (*D*) of the *S*s is maintained when the extended CS-UCS interval is employed, whereas in the traditional procedure it decreases because the UCS is discontinued.

However, a subsequent study by Reynolds (1958) has cast some doubt as to the adequacy of this interpretation. While one part of Reynolds' experiment supported the findings of McAllister, another portion did not. McAllister's groups, it should be noted, had been conditioned with a partial (80%) reinforcement schedule. Confirming McAllister's results, Reynolds found that when the original conditioning involved a partial reinforcement schedule (60%), the group with the extended CS-UCS interval extinguished much more slowly than the group in which the UCS was omitted. In the case of *S*s conditioned originally with continuous (100%) reinforcement, however, Reynolds failed to find a significant difference in performance under the two extinction procedures, the extended CS-UCS interval group showing the same precipitous decrease in response strength characteristic of extinction with the UCS omitted. It would appear from these findings that the difference in extinction performance under the two procedures is contingent upon the original conditioning being conducted under a partial reinforcement schedule.

The present study is the first of a series aimed at furthering our understanding of extinction of the classical aversive CR in humans. The purpose of the experiment here reported was first, to ascertain the generality of the findings of Reynolds following continuous reinforcement under the above extinction conditions, particularly whether the same results would hold when, in contrast to his study, no ready signal is employed. Secondly, we were interested in the interaction of different levels of original conditioning with the two procedures on the extinction data, and finally, we were desirous of looking at the course of extinction in smaller trial blocks rather than in terms of large blocks of trials. In view of the rapidity of extinction of the eyelid CR previously reported studies have presented the data in much too large trial blocks.

---

[1] This study was carried out as part of a project concerned with the influence of motivation on performance in learning under Contract Nonr-1509(04), Project NR 154-107 between the State University of Iowa and the Office of Naval Research.

METHOD

Subjects.—A total of 132 men from an introductory course in psychology was employed. Of these, 15 were eliminated for having given more than 50% responses of the form used to define voluntary responders (cf. Spence & Ross, 1959). The data of 17 other *S*s were also discarded, 2 for giving CRs on two of the pretest trials, and 15 for giving less than five CRs during the first 32 acquisition trials. The remaining 100 *S*s provided four groups, each of which contained 25 *S*s.

Apparatus and method of recording.—The *S* was seated in a dental chair in a semidarkened soundproof room. The *E* was in an adjoining room in which the stimulus controls and recording equipment were located. The apparatus for presenting the stimuli and recording the eyeblinks was identical with that used in previous studies from this laboratory (cf. Spence, 1953).

The CS consisted of a 500-cps tone with an intensity of 70 db. (re .0002 dynes per cm^2). During acquisition the duration of the CS was 550 msec., with the UCS occurring 500 msec. after the onset of the CS. During extinction two conditions were used. In one condition the CS was presented for 2,550 msec., with the UCS occurring 2,500 msec. after the onset of the CS. In the other, the CS was presented for 550 msec. and the UCS was omitted. The duration of the UCS, an air puff of 2 psi applied to the right eye through a .062-in. diameter orifice, was limited to 50 msec. by a 115-v., 60-cycle ac-operated solenoid valve controlled by an electronic timer.

The *S* was instructed to fixate upon and keep looking at a 6-cm. milk glass disc located directly in front of him. He was also instructed to be as relaxed as possible, and not to attempt to control the responses of his eye. A CR was recorded whenever the record showed a deflection of 1 mm. or more in the range 150–500 msec. following the onset of the CS. Responses in the interval 0–150 msec., which were infrequent, were classified as original responses to the CS and not included in the data.

Conditioning procedure.—Following the reading of the instructions, each *S* was given three presentations of the CS alone followed by a single presentation of the 2-lb. UCS alone. Each *S* then received 32 or 64 conditioning trials which involved the paired presentation of the CS and UCS with a 500-msec. CS-UCS interval. Intertrial intervals of 15, 20, and 25 sec., averaging 20 sec. and arranged according to a fixed schedule, were used throughout the experiment.

Experimental design.—The experiment consisted of two phases. In Phase 1, which lasted 32 trials for Groups IA and IB and 64 trials for Groups IIA and IIB, all *S*s received standard conditioning trials. Immediately following the last conditioning trial, with no break in the continuity of the experiment, Phase 2 was begun. In this phase Groups IA and IIA received 40 extinction trials with the UCS omitted, while Groups IB and IIB received the same number of extinction trials with the UCS following the CS by 2,500 msec.

RESULTS

Frequency curves of extinction by blocks of 2 trials over the first 10 trials and by blocks of 10 trials over the subsequent 20 trials are shown in Fig. 1. These curves differ from extinction curves of eyelid responses that have been presented heretofore in that the abscissa is in terms of the number of *previous* nonreinforced trials. Under this procedure the data for the first point (0) on the graph, which is the first trial on which no reinforcement is given, reflects performance *after* the last reinforced trial (i.e., after 32 or 64 reinforcements),

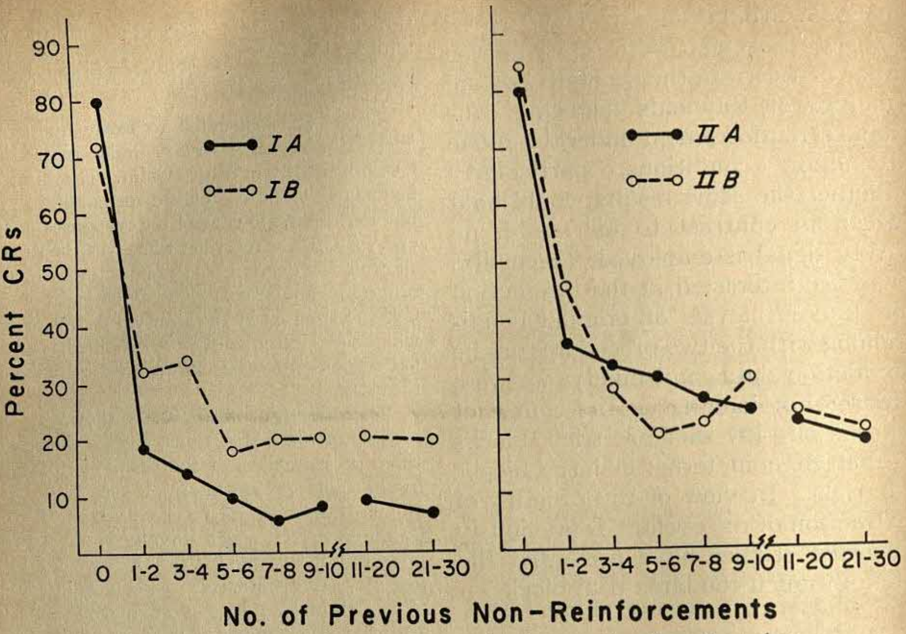


FIG. 1. Percentage of CRs during extinction as a function of number of previous nonreinforced trials.

while the data for Trials 1–2, 3–4 etc., represent the average frequency of response to the CS *after* the indicated number of nonreinforced trials.

One of the most striking features of these group extinction curves is the extremely rapid rate with which response strength decreases with the shift to nonreinforcement. As may be seen in the case of three of the groups, the average performance was down to a minimal level after only 5–8 nonreinforcements. Examination of the data of individual *S*s reveals that for the majority of *S*s extinction was even much faster than suggested by these group curves. The frequency distribution of individual data (number of CRs in 30 trials) was J shaped ranging from 0 to 24 CRs. The extent to which the scores piled up at the lower end of the distribution is shown by the number of *S*s (21), who never gave a single CR and by the fact that a majority of the *S*s (53%) gave three or less CRs. Another indication of the speed with which extinction occurred is the finding that after the third nonreinforced trial 60% of the *S*s responded at a level of one CR or less in the ensuing 10 trials. The precipitous nature of this decremental effect of nonreinforcement has, heretofore, been obscured by the inclusion of the responses made on the first nonreinforced trial and the use of relatively large blocks of trials (5, 6, or 10).

TABLE 1

NUMBER OF CRs OCCURRING IN EXTINCTION

Groups	Trials 1–10		Trials 1–30	
	M	*SD*	*M*	*SD*
IA	1.12	1.58	2.88	3.68
IB	2.48	2.71	6.52	8.24
IIA	2.96	2.63	6.80	8.90
IIB	2.92	2.91	7.24	7.78

Comparing next the extinction performances of the different groups, Fig. 1 shows that the curve of Group

IA tends to fall below those of the other three groups, all of which appeared to perform at approximately the same level. Table 1 presents the data for the four groups in terms of the mean number of CRs given in Trials 1–10 and 1–20. Statistical evaluation by means of the Mann-Whitney *U* test of the differences in number of responses made in these extinction periods by the groups (I and II) given different amounts of original training revealed the difference for Trials 1–30 to be significant (two-tailed test) at the .05 level ($z = 1.98$) while that for Trials 1–10 was reached at the .09 level ($z = 1.69$). The differences between *S*s (A and B) extinguished under different conditions (no UCS and extended CS-UCS interval) were not significant, the z values being .85 for Trials 1–10 and 1.48 for Trials 1–30.

Since the data suggested the possibility of an interaction between the two experimental variables, with the extinction conditions possibly having a differential effect at the lower training level but not at the higher, a separate comparison (Mann-Whitney *U*) was made of the performances of the two low trained groups (IA and IB). This difference, however, was also not significant ($z = 1.38$).

Discussion

In failing to find evidence of a difference in performance level under the two extinction procedures following continuous reinforcement the results of the present study, in which a ready signal was not employed, confirm the finding of Reynolds (1958), who did employ a ready signal. This finding precludes the interpretation of McAllister of his positive results following partial reinforcement as being due to the maintenance of a higher drive (*D*) level under the extended CS-UCS interval than when the UCS is omitted. Such a hypothesis would require the same finding following a continuous reinforcement schedule.

As was pointed out in the introduction, the available evidence indicates that the finding of a difference under the two extinction procedures (extended CS-UCS interval or no UCS) appears to be dependent on the use of a partial reinforcement schedule in the original acquisition of the CR. One possible interpretation of these phenomena would be in terms of some form of the discrimination hypothesis that has been put forward to explain the so-called Partial Reinforcement Effect (PRE), i.e., slower extinction of a CR following partial than after continuous reinforcement. While the present authors favor a conditioned frustration interpretation of PRE in the case of learned behavior involving appetitional needs and reward incentives (Amsel, 1958; Spence, 1960), we believe that extinction in classical aversive conditioning is a very different phenomenon involving a quite different mechanism.

The particular version of the discrimination hypothesis that we at present favor involves the somewhat neglected concept of set (attitude, expectancy). In the early work on conditioning in human *S*s (Hilgard & Marquis, 1940) considerable attention was given to this hypothetical concept, which was variously defined in terms of instructions from *E* or self-instructions by *S* himself. A number of investigators (c.f., Grant, 1939; Hilgard, Campbell, & Sears, 1938; Norris & Grant, 1948) have clearly shown that variables defined in terms of verbal instructions administered by *E* can have important effects on the level of performance in simple conditioning and on the degree of discrimination in differential conditioning. The important role that self-instructions may play in conditioning is evidenced by those *S*s who give few or no CRs even after 100 conditioning trials, or at the opposite pole, *S*s who, almost from the start of conditioning, consistently respond voluntarily to the CS prior to the onset of the UCS.

The standard instructions employed in a conditioning experiment attempt to minimize the role of such cognitive sets or attitudes by requesting *S* not to attempt to control the response of his eye, to think of other things, etc. To a considerable degree these instructions have been found to be successful as far as the acquisition period is concerned, for most *S*s appear to follow the instructions and their verbal reports even indicate that they do not recognize the nature of the experiment. However, we have found that all *S*s report that they recognize the change in the experimental procedure when the UCS is omitted for the first time as is the case following continuous reinforcement in the acquisition period. The rapidity with which extinction occurs under this condition suggests that with the cessation of the UCS most *S*s almost immediately adopt a strong inhibitory set. In the case of extinction following intermittent reinforcement, however, this inhibitory set apparently does not develop, at least not immediately. Under this reinforcement schedule *S* has already experienced failures of occurrence of the UCS on some portion of the trials, hence, the shift to 0% reinforcement is not readily recognizable as in the case of the shift from 100% to 0%. Performance in the early stages of extinction following a partial schedule of reinforcement would thus not be affected by such an inhibitory set and, hence, would not be expected to show the precipitous drop that occurs following continuous reinforcement.

According to this interpretation, an important factor determining the development of such an inhibitory set with the institution of extinction is the degree to which the acquisition and extinction conditions are readily discriminable by *S*. In this connection, it is interesting to note that, whereas earlier studies of eyelid conditioning (Grant & Hake, 1951; Grant & Schipper, 1952; Grant, Schipper, & Ross, 1952; Humphreys, 1939) had uniformly shown that performance decreased more slowly following partial than after continuous reinforcement, Reynolds failed in the case of *S*s extinguished without the UCS to find such a difference. Instead, both groups showed a rapid decline to a random level of response (10%). A major difference in Reynolds' experiment was that the shift to extinction involved a much greater change in the conditions for his partial group than had been the case in the earlier studies. Whereas, the latter had used no UCS on the nonreinforced trials during acquisition, Reynolds employed the extended CS-UCS interval. The shift from the long CS-UCS interval on a portion of the trials to no UCS on any trial was, of course, much more discriminable than was the shift in the earlier studies. On the other hand, Reynolds did find that partial reinforcement led subsequently to slower extinction than continuous reinforcement under the condition in which the extended CS-UCS interval was employed during extinction. This may be seen to be in line with our set interpretation. That is, shift from an acquisition schedule that involved experience on 40% of the trials with the long CS-UCS interval to one in which all of the trials had the long CS-UCS interval would be difficult to discriminate, whereas a shift from continuous reinforcement with all short CS-UCS intervals to all long interstimulus intervals would be readily discernible.

A similar account in terms of this situation may be given of Reynolds' finding that, following conditioning with partial reinforcement, extinction is slower when the extended CS-UCS interval is employed than when the UCS is omitted. However, failure to obtain a difference under these two extinction procedures following conditioning with a continuous reinforcement schedule, as both our study and Reynolds found, is somewhat embarrassing to the present interpretation. Since a shift from 100% reinforcement to the extinction procedure involving the long CS-UCS interval is less different from that in which the shift is to no UCS, one might have expected the former condition to have led to somewhat greater resistance to extinction.

Questioning of *S*s who were extinguished under the extended CS-UCS procedure following continuous reinforcement in a recent doctoral study by Goldstein (1962) revealed that all *S*s readily recognized the change at the beginning of the extinction. This recognition was undoubtedly facilitated in all of these experiments by the fact that the shift to extinction also involved lengthening the CS duration from 550 msec. to 2,550 msec. as well as lengthening the interval between the onset of the CS and UCS.

Further studies are needed in which an attempt is made to employ procedures in which the shift to extinction following continuous reinforcement is less discriminable. Thus, the technique employed in the present study could be made less recognizably different from the reinforcement conditions by having the duration of the CS 2,550 msec. during acquisition as well as during extinction, with only the interstimulus interval being extended. Also needed are studies in which there is less likelihood that the human *S* will be aware that either conditioning or extinction is occurring. A number of studies of this kind are in progress in our laboratory. The technique has involved embedding the conditioning in the context of a light-guessing experiment. Under these conditions none of the *S*s have reported recognizing that they were being conditioned, nor have they been aware of the change in the conditions from acquisition to extinction. Most significant is the finding that extinction has been extremely slow and gradual in these experiments, the large initial drop in performance being entirely absent.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55,** 102–119.

Goldstein, H. The effects of set factors and reinforcement schedules on the acquisition and extinction of the conditioned eyelid response. Unpublished doctoral dissertation, State University of Iowa, 1962.

Grant, D. A. The influence of attitude on the conditioned eyelid response. *J. exp. Psychol.*, 1939, **25,** 333–346.

Grant D. A., & Hake, H. W. Dark adaptation and the Humphreys random reinforcement phenomenon in human eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 417–423.

Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43,** 313–320.

Grant, D. A., Schipper, L. M., & Ross, B. M. Effects of inter-trial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44,** 203–210.

Hilgard, E. R., Campbell, R. R., & Sears, W. N. Conditioned discrimination: The effect of knowledge of stimulus-relationships. *Amer. J. Psychol.*, 1938, **51,** 498–506.

Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25,** 141–158.

McAllister, W. R. The effect on eyelid conditioning of shifting the CS-UCS interval. *J. exp. Psychol.*, 1953, **45,** 423–428. (a)

McAllister, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45,** 417–422. (b)

Norris, E. B., & Grant, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *Amer. J. Psychol.*, 1948, **61,** 37–49.

Reynolds, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55,** 335–341.

Spence, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45,** 57–63.

Spence, K. W. *Behavioral theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Spence, K. W., & Ross, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58,** 376–381.

(Received October 1, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 292–299

EFFECTS OF "ANXIETY-LESSENING" INSTRUCTIONS AND DIFFERENTIAL SET DEVELOPMENT ON THE EXTINCTION OF GSR[1]

WILLIAM W. GRINGS AND RUSSELL A. LOCKHART

University of Southern California

Results of previous investigators indicate a failure of "anxiety-lessening" instructions to extinguish a GSR conditioned under long (e.g., 6 sec.) CS-UCS interval conditions. This study observed the effects of such instruction on GSR extinction under conditions varying UCS intensity and number of reinforcements. Results: (a) GSR extinguished more rapidly and to a greater extent under conditions of anxiety-lessening instructions than under conditions of noninstruction, (b) the instruction effect was independent of number of reinforcements and UCS intensity, (c) increasing GSR magnitude was observed during extinction in Ss previously giving large disparity responses—even when administered anxiety-lessening instruction—emphasizing the importance of assessing S's reaction to changed conditions as a determiner of subsequent responding.

It has been a common finding that instructional parameters affect the course of conditioning human *S*s. A classic example is that of Cook and Harris (1937) in which the magnitude of GSR increased when *S* was told he would receive a shock, and decreased when told he would not receive a shock. Several direct attempts have been made to understand the role of instructional set on both acquisition and extinction of the CR. These studies have been of two general sorts: (*a*) those which instruct *S* on how to respond (Hilgard & Humphreys, 1938; McAllister & McAllister, 1958; Miller, 1939; Norris & Grant, 1948), and (*b*) those which instruct *S* primarily concerning the stimulus situation (Cook & Harris, 1937; Lindley & Moyer, 1961; Mowrer, 1938; Silverman, 1960).

The general result indicates that response frequency and/or magnitude can be reliably manipulated through instructional set. A typical explanation for such effects can be seen in the recent work of Lindley and Moyer (1961). They predicted the effects of instructions on extinction of the finger withdrawal response in terms of S-R drive concepts. If "no more shock" instructions reduce general drive level, then by virtue of the multiplicative relation between drive and habit strength, the degree of observed responding should decrease. Moreover, if the effect of instruction is constant across habit strengths, then instruction should be less effective when the habit is strong than when it is weak. These predictions have been confirmed for conditioned skeletal responses.

A recent result suggests that instructional effects are more complicated for conditioned autonomic responses. Silverman (1960) found that a group conditioned under a 6-sec. CS-UCS interval and administered anxiety-reducing instructions ("No more shock") showed the same decrement in magnitude of GSR during extinction as a similarly conditioned, but noninstructed group. Silverman was led to postulate that

---

[1] Supported by a research grant (M-3916) from the National Institute of Mental Health.

in long delay conditioning an anxiety response is developed which is not affected by "no more shock" instructions, and functions to maintain GSR responding during extinction.

In light of this atypical result and interesting hypothesis, the present study was designed as a further test of the effects of anxiety-reducing instructions on extinction of the GSR in long delay conditioning with variation in number of reinforcements (CS-UCS pairings) and strength of reinforcement (UCS intensity).

METHOD

Subjects.[2]—Forty paid undergraduate volunteers served as *S*s. All *S*s were found to have normal hearing and normal color vision after audiometric examination and administration of the Ishihara Color Test prior to experimentation.

Apparatus.—Three small colored geometric forms served as separate CSs. These were produced by projection through filters onto a ground glass screen located in front of *S*. The three colored forms used (red triangle, blue square, and yellow circle) were obtained from standard Wratten filters (47, 12, and 28, respectively) and were matched for brightness and area. Three intensities of UCS (shock) were employed. The shock was a condenser discharge through a transformer and series resistor to a point on *S*'s right forearm. The CS duration and the interstimulus interval were controlled by Hunter electronic timers. The CS duration was 5 sec., terminating with the instantaneous condenser discharge. Intertrial intervals were manually varied in an irregular fashion among 15, 20, and 25 sec. The GSR was obtained as a dc resistance change through $\frac{1}{2} \times \frac{5}{8}$ in. silver electrodes bent to the contour of *S*'s first and third fingers of the left hand, coated with electrode paste, and taped firmly to the finger tip. Recording of the GSR was accomplished by a Darrow-type bridge input to a two-channel dc amplifier (Brush, Model 530) which drove an Esterline Angus ink writing milliammeter.

Design.—Three independent variables were studied: (*a*) *Number of reinforcements*—half the *S*s were given 9 and half 36 CS-UCS pairings (3 or 12 on each of 3 CS-UCS pairs); (*b*) *Intensity of reinforcement*—all *S*s received three intensities (high, medium, and low) of UCS each consistently paired with a particular CS; (*c*) *Instruction*—half the *S*s received, prior to the extinction series, instructions to the effect that no further shocks would be received, while the other half received no such instruction.

The *S*s were assigned to the independent variable conditions by a modified random method by *S* number prior to the experimentation. Assignment of cues to intensity levels was performed by systematic rotation. The order of presentation of stimuli during acquisition was set for each *S* by using random blocks of nine with the restriction that not more than two successive presentations of the same CS-UCS pair be allowed.

Procedure.—The *S* was brought into a soundproof room and seated at a table in front of a translucent screen embedded in a partition which separated *S*'s quarters from *E* and the recording and stimulus-production equipment. The *S* was told that shock was to be administered during the course of the experiment, the intensity of which he would set himself. (Two *S*s refused to submit to shock and were replaced.) The GSR and plethysmograph pickup devices were then attached to *S* and the function of each explained. The shock electrodes (two silver discs attached to a leather strap) were then placed on the volar surface of the right arm. The *S* received a copy of typewritten material including a discussion of the nature of the GSR (e.g., "sweat gland activity of the finger tip," etc.), a general orientation to the situation (e.g., "keep your hands and arms as still as possible and your eyes on the screen," "you may stop at any time," etc.) and instructions to guess out loud prior to CS termination which level of shock intensity will follow.

A "work-up" method was used in determining *S*'s threshold and higher intensity settings for shock. The *S* was told to report when the shock was first perceived. This shock was called Level 1. The *E* then increased the shock intensity gradually until a point was reached where *S* reported discomfort, unwillingness to proceed to the next higher intensity, or until the maximum output of the stimulator had been reached. This point was called Level 3. A point was then determined by *E* to represent the middle stimulus (Level 2) which was a voltage applied to the condenser two-thirds of the distance between voltages for Levels 1 and 3.

---

[2] The authors are indebted to Sidney Carlin for his assistance in collecting of data.

TABLE 1

MEAN MAGNITUDE OF GSR DURING THREE EXTINCTION TRIALS IN COMPARISON WITH LAST ACQUISITION TRIALS FOR REINFORCEMENT AND INSTRUCTION GROUPS

Group	Short Reinforcement			Long Reinforcement		
	Acquisition	Extinction	Diff.	Acquisition	Extinction	Diff.
Instructed	.93	.53	.40**	.54	.24	.30**
Noninstructed	.76	.88	−.12	.85	.71	.19

Note.—First interval response only.
** $p < .01$.

The mean voltage levels produced in this manner for the instructed group were 27.4, 62.1, and 85.3; and 29.5, 69.3, and 95.3 for the noninstructed group. For each level the difference between these means is not significant ($p > .20$ in each case). Likewise, there were no differences in variability on the low ($\sigma = 12.0$ and 13.3), medium ($\sigma = 17.3$ and 20.3), or high ($\sigma = 23.9$ and 29.6) shock cues for the instructed and noninstructed groups respectively.

Following the reinforcement series, two identical disparity trials were given, these consisting of a cue followed by a shock intensity different from the shock intensity previously experienced with that cue. After disparity, two further training trials were given. The first consisted of the cue used

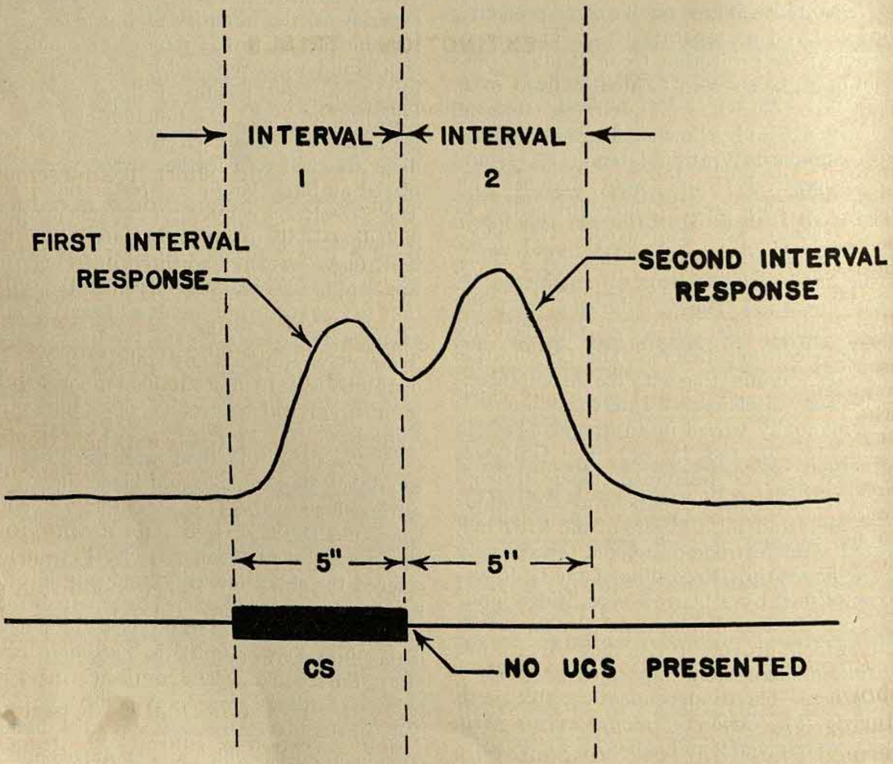


FIG. 1. Characteristic GSRs in first and second interval periods of extinction trials where the UCS is withheld.

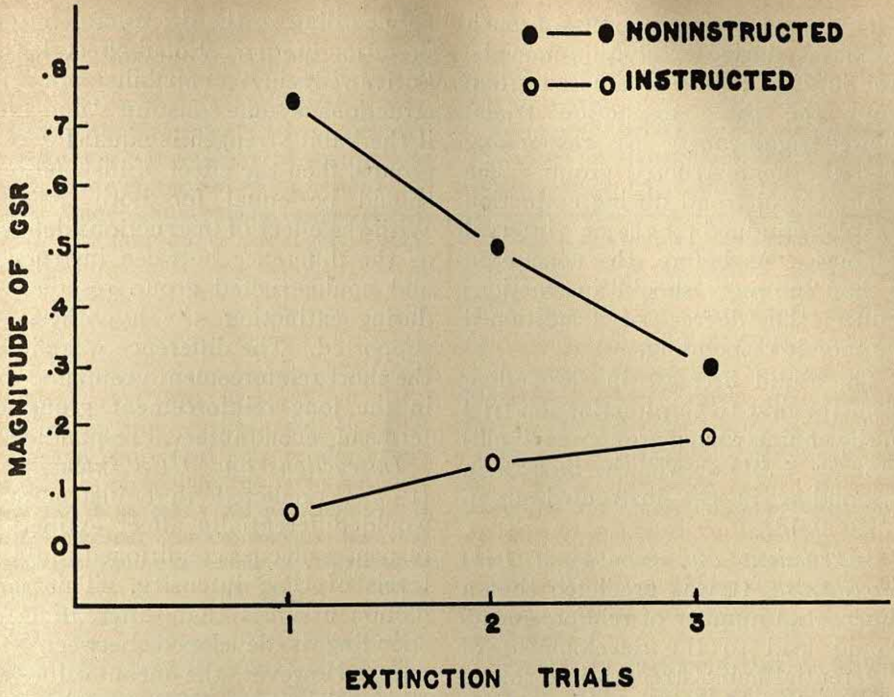


FIG. 2. Magnitude of second interval GSR during extinction for instructed and noninstructed groups.

in the disparity trials followed by its "proper" (i.e., previous to disparity) UCS intensity. The second consisted of the cue appropriate to the level of UCS intensity used in the disparity trials followed by its "proper" UCS.

Half the *S*s were told after the last training trial that, "There will be no more shock; there will be no more shock." After these instructions, each CS was presented once (in a randomized order) without reinforcement with each *S* receiving a random permutation of three trials. The other *S*s received no instructions prior to the extinction period. Following extinction, *S*s were interviewed concerning the experiment and were told not to tell anyone of their experiences in the laboratory.

RESULTS

Effect of instructions.—Table 1 shows the mean GSR magnitude during the delay period (hereafter termed "first interval" response) for the acquisition (last response on each cue) and extinction periods. For both long and short reinforcement groups, GSR magnitude was reduced significantly following instructions. Changes in the noninstructed group were not significant. Thus, so-called "anxiety-lessening" instructions are effective in reducing anticipatory GSR responding in the long delay conditioning situation.

Because the UCS is withheld during extinction, it was possible to measure the *second interval response.* This response, described in a previous study (Grings, Lockhart, & Dameron, 1962), occurs in delay conditioning on test or extinction trials when the normally presented UCS is withheld. It occurs only *after* the delay interval (i.e., during the normal UCR period). Such a response, and its relation to the anticipatory GSR is illustrated in Fig. 1.

The effect of instructions on such responses during extinction can be seen in Fig. 2. The instructed group (over the three extinction trials) showed significantly less responding than the noninstructed group. The same general trend during extinction was also obtained for the first interval response. As before, the conclusion is that "no more shock" instructions reduce the degree of conditioned autonomic responding.

The rapid *decrease* in responding from the first to third extinction trial in the noninstructed group was significant ($p < .01$) as was the *increase* in responding in the instructed group ($p < .03$).[3]

Instructions and number of reinforcements.—It was predicted that a differential number of reinforcements would lead to the development of differential habit strength in terms of expectation of UCS. To test for habit strength differences, the *disparity response*[4] data were analyzed. This response has been used previously to measure the strength of set development (Grings, 1960). However, when the short and long reinforcement groups were compared, no significant difference was found in degree of perceptual disparity responding. To the extent that such a measure is sensitive to differential set development it can be concluded that the set to expect UCS was about equally conditioned in both reinforcement groups.

According to the drive-habit interpretation of instructional effect, if the "drive-reducing" capability of instructions is some constant value, and if the habit strength is equal for two groups, then the effect of instructions should be equal for both groups. With the effect of instructions defined as the difference between instructed and noninstructed group responding during extinction, the hypothesis is supported. The difference was .50 in the short reinforcement group and .52 in the long reinforcement group in terms of second interval responding.

Instructions and UCS intensity.—It was predicted that instructions would differentially affect extinction responses on cues conditioned to three levels of UCS intensity. This prediction assumes that differential responding is developed between the cues. However, the mean values of the anticipatory GSR on the last acquisition trial (across reinforcement conditions) for each cue were not significantly different (.76, .67, and .84 for the low, medium, and high shock, respectively). If these values reflect an equal set strength, the drive-habit interpretation would again predict a constant decrease in responding following instructions. The differences in anticipatory responding from acquisition to extinction were .39, .34, and .44 for the low, medium, and high expected shock cues, respectively. Since these values were not significantly different, the drive-habit interpretation gains further support.

DISCUSSION

Effect of instructions.—Instructions were effective in reducing GSR magnitude from acquisition to extinction. More importantly, anxiety-reducing instructions resulted in significantly less responding than that observed in a comparable noninstructed group. It is apparent that instructions *facilitated* the extinction of the GSR, a result in keeping with most experimental studies in this area. When the present study is compared with Silverman's, one conclusion is suggested: the failure of instructions to facilitate extinction of the conditioned GSR under Silverman's experimental conditions may not be the result of some process inherent in long delay autonomic conditioning. If this were the case, the present results would not have been obtained.

---

[3] All analyses were nonparametric. Between-group comparisons were carried out by the Mann-Whitney *U* test, and within-group comparisons by the Wilcoxon paired replicates test.

[4] Disparity trials consist of following a cue with a shock differing in intensity from that used consistently throughout training. Response to this change in conditions (corrected for absolute stimulus intensity) is termed the perceptual disparity response.

Differential reinforcement.—The data indicate that the strength of UCS expectancy was equally conditioned for both the short and long reinforcement groups, even though the latter received four times as many reinforcements. It was therefore impossible to test the assertion that the effect of instructions would vary as a function of differing numbers of reinforcements. However, instructions produced a constant decrement (in absolute value) in responding for two groups receiving different numbers of reinforcements, a result which is consistent with a drive-habit interpretation of instructional effect.

Differential UCS intensity.—Anticipatory GSR responding was not significantly different among cues representing three levels of UCS intensity on the last trial prior to extinction (in either reinforcement group). Therefore, it was not possible to adequately test the hypothesis that instructions would interact differentially with the expected UCS intensity. However, an approximately equal decrement (in absolute value) was observed from acquisition to extinction in anticipatory responding to each cue. This result is also consistent with a drive-habit interpretation. Significant discrimination between the cues did not develop in the short reinforcement group, but did develop in the long reinforcement group after six reinforcements on each cue. However, this performance difference disappeared with continued reinforcement. Thus, discrimination was established, but disappeared after a few reinforcements. This effect may be related to a common observation in GSR conditioning, i.e., decreasing anticipatory responding with continued reinforcement after reaching a peak between 6–8 reinforcements.

Increasing response during extinction.—The effect of instructions was to significantly reduce responding from acquisition to extinction. However, superimposed on this effect was a further effect describable as a significant *increase* in magnitude of response from the first to third extinction trials for both anticipatory and second interval GSRs.

The reason for such an increase is not immediately clear. Again, it is possible that the introduction during acquisition of disparity trials could have contributed determining variation. On previous occasions it has been found that some *S*s respond markedly in disparity situations, while other *S*s appear to be unaffected by the change in conditions and show small or negative disparity responses.

It might be predicted that *S*s giving large disparity responses would also anticipate future change in conditions more so than small disparity responders. That is, persons responsive to changed conditions might be "sensitized" to anticipate further changes in conditions.

Figure 3 shows the extinction performance to *S*s divided on the first disparity trial into above median and below median disparity responders. It can be seen that in terms of second interval responding, persons in the instructed group giving above median disparity responses are the ones who respond at an increasing rate during extinction—even when told that no more shocks were to be received.[5]

A question may be raised as to the influence of instructions on this increase. That is, are instructions of no more shock necessary to produce differential responding during extinction between

---

[5] Because such responding might represent a personality factor, *S*s in the two disparity groups were compared on all scales of the MMPI and the Guilford-Zimmerman Temperament Survey (GZ). The only significant difference was found on the S scale (sociability) of the GZ.

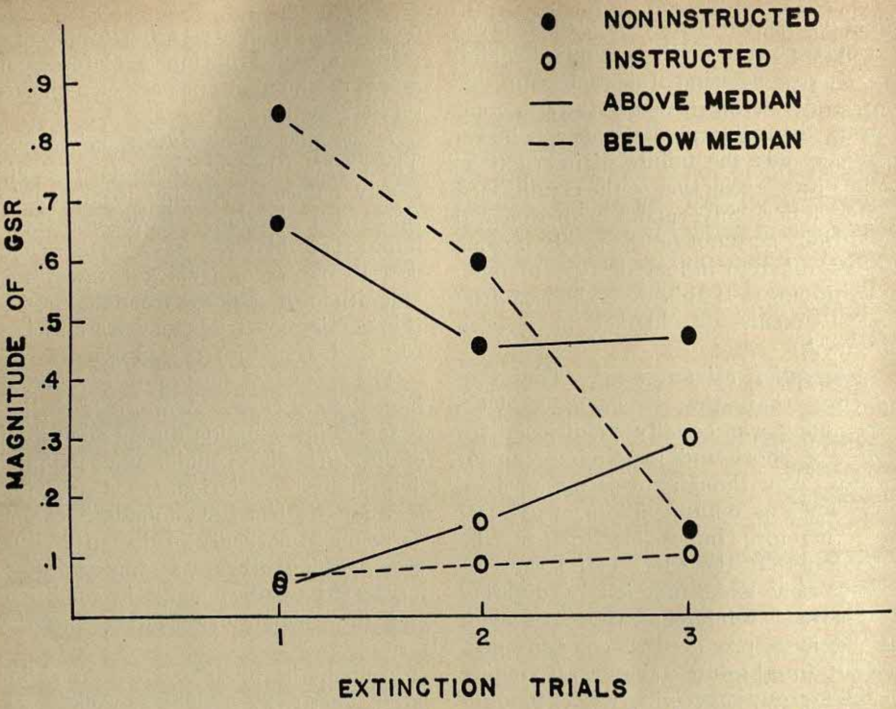


FIG. 3. Magnitude of second interval GSR during extinction for instructed and noninstructed *S*s giving above median and below median disparity responses in a prior disparity situation.

small and large disparity responders? In order to assess this possibility, *S*s in the noninstructed group were divided into above and below median disparity responders.

It can be seen in Fig. 3 that the below median disparity responders in the noninstructed group show a more rapid decrement in second interval responding during the extinction period than do the above median disparity responders. The overall difference between the two groups (comparing a difference of −.20 and −.70) is significant beyond the .01 level.

It would appear then that *S*s who overly respond to changed conditions develop a persisting expectation of further change in conditions. Overt instructions given by *E* (and designed to objectively reduce the threat of shock) or the occurrence of an unannounced extinction series actually enhance and amplify this anticipated disconfirmation.

This result indicates the possible use of the disparity situation as an independent variable in studies of preparatory perceptual set. In addition, this result emphasizes the importance of assessing *S*'s reaction to changed conditions as a determiner of subsequent responding.

REFERENCES

COOK, S. W., & HARRIS, R. E. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, 21, 202–210.

GRINGS, W. W. Preparatory set variables in the classical conditioning of autonomic variables. *Psychol. Rev.*, 1960, **67**, 243–252.

GRINGS, W. W., LOCKHART, R. A., & DAMERON, L. E. Conditioned autonomic responses of mentally subnormal individuals. *Psychol. Monogr.*, 1962, **76**(39, Whole No. 558).

HILGARD, E. R., & HUMPHREYS, L. G. The effect of supporting and antagonistic voluntary instructions on conditioned discrimination. *J. exp. Psychol.*, 1938, **22**, 291–304.

Lindley, R. H., & Moyer, K. E. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *J. exp. Psychol.*, 1961, **61**, 82–88.

McAllister, W. R., & McAllister, D. E. Effect of knowledge of conditioning upon eyelid conditioning. *J. exp. Psychol.*, 1958, **55**, 579–583.

Miller, J. The effect of facilitory and inhibitory attitudes on eyelid conditioning. Unpublished doctoral dissertation, Yale University, 1939. Cited by E. R. Hilgard & D. G. Marquis, *Conditioning and learning.* New York: Appleton-Century, 1940.

Mowrer, O. H. Preparatory set (expectancy): A determinant in motivation and learning. *Psychol. Rev.*, 1938, **45**, 62–91.

Norris, E. B., & Grant, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *Amer. J. Psychol.*, 1948, **61**, 37–49.

Silverman, R. E. Eliminating a conditioned GSR by the reduction of experimental anxiety. *J. exp. Psychol.*, 1960, **59**, 122–125.

(Received October 5, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 300–307

REPRODUCTION OF BINARY VISUAL PATTERNS HAVING DIFFERENT ELEMENT-PRESENTATION SEQUENCES[1]

E. RAE HARCUM AND STEPHEN M. FRIEDMAN

College of William and Mary

27 Os reproduced binary patterns with the elements presented in 8 spatial-temporal conditions, involving right-to-left, left-to-right, or random sequences, or presentation of all elements simultaneously. Fewer errors appeared to the left of fixation relative to the right, but only when all 10 elements appeared simultaneously. Simultaneous exposure produced fewer errors than did sequential exposure. No large differences in totals of errors among the sequential-presentation conditions appeared, but "recency" effects, appearing in the right-to-left and left-to-right sequences, were stronger than "primacy" effects. Orderly sequential presentation of the elements produced superior performance for individual Os when the sequence proceeded from the end showing more errors in the control, random-presentation condition.

With certain visual patterns composed of elements flashed simultaneously on both sides of the fixation point, generally greater perceptual accuracy occurs for those elements within the pattern which appear to the left of fixation (Anderson, 1946; Bryden, 1960; Crosland, 1939; Harcum, 1957, 1958; Harcum & Dyer, 1962; Harcum & Rabe, 1958; Heron, 1957; Terrace, 1959). This greater accuracy at the left has been attributed to a primacy effect in a sequential perceptual process which operates like a scanning of the eyes across a visual stimulus in reading (Bryden, 1960; Harcum, 1957, 1958; Heron, 1957; Terrace, 1959). If this hypothesis is correct, some temporal-spatial presentation sequences of elements should produce different perceptual performance than is produced by other sequences.

Mackworth (1962) has shown that more accurate visual perception occurs when all of the digits in a pattern are presented simultaneously than when they are presented in sequence. Furthermore, a left-to-right sequence of presenting the elements might be superior to a right-to-left sequence for English reading *O*s, because eye movements from left to right across the visual field during reading are superior for Western readers (Aulhorn, 1948; Tinker, 1955). Although Foley (1957) found superior perception for digits moving from right to left, which he accounted for on the basis of the inverted retinal image, his visual task was quite different from the present one. Our argument is that the motor components of the scanning process cause the traces of the individual elements to be perceived in the same sequence that the actual target elements would be fixated by the eyes in reading. Recent evidence for this is that a direct relationship exists between the side of fixation favored by perceptual accuracy in tachistoscopic exposure and the side fixated by the first postexposure eye movement (Bryden, 1961; Crovitz & Daves, 1962).

The present study determines the

---

[1] This investigation was supported by Public Health Service research grants (B-1971 and B-2661) from the Institute of Neurological Diseases and Blindness. These results have been previously reported to the annual meeting of the Virginia Academy of Science (Friedman, 1961).

effects on the reproduction of binary visual patterns by presenting the stimulus elements horizontally across the visual field in eight different temporal-spatial conditions. The following hypotheses are tested: (*a*) a binary pattern exposed tachistoscopically across the visual field is sequentially scanned by a perceptual mechanism in a manner related to the reading history of *O*s, such that, for English reading *O*s, primacy effects produce fewer errors for the elements at the left; (*b*) scanning of the visual field proceeds in an orderly fashion across the visual field so that orderly temporal-spatial presentation of the elements produces more accurate perception than a random order of presentation; and (*c*) when the presentation sequence corresponds to the dominant temporal-spatial characteristics of the scanning process of *O*, the subsequent perceptual accuracy is greater. In other words, orderly sequential presentation of the elements will produce superior performance, especially when the sequence proceeds from the end of the pattern which normally exhibits the fewer errors under control conditions.

METHOD

Apparatus.—The apparatus was a Perceptoscope Mark 5102-2 projector,[2] set to expose individual frames of 16-mm. film for .125 sec. Stimulus elements were produced on the film by photographing black outline circles on white paper—5 of the 10 circles for each target having been filled in with black ink to produce binary-element patterns. The projected image of each circular element was about 15.4′ in diameter; the total target was about 4.3° long, viewed from a distance of 14 ft. The target was exposed along the horizontal diameter of a projected circular white field, which was about 10.3° in diameter. The exact dimensions of the projected circles are difficult to specify because of the graduated edge of a projected contour. The size and contrast of the filled circles appeared to be sufficient to make them detectable if presented singly, but the stroke-width and contrast of the open circles probably were not sufficient to make them always clearly suprathreshold.

A projected fixation cross appeared in the center of the projected circular white field when the target was not present. This cross registered with the center of the target when it appeared.

Observers.—The 27 *O*s, all college students with normal vision, were used in groups of from 2 to 6 at a time, seated at tables arranged in a semicircle in front of the projection screen. The field luminance of 8.2 ft-L was about the same from each *O* station.

Procedure.—The presentation of each target under each of eight temporal-spatial conditions required the exposure of 11 frames of the film. The presentation sequence for a single observation consisted of first presenting a fixation cross for an indefinitely long exposure until all *O*s were ready for the stimulus exposure. The *E* then said, "Ready," and started the projector. Eleven frames of the film were immediately exposed in sequence at the rate of 8/sec. The first 10 frames contained the stimulus pattern under the appropriate one of eight experimental conditions. The eleventh frame which followed contained the fixation cross again; it remained visible until the next exposure sequence was initiated. The *O*s were instructed to maintain fixation during the entire target exposure interval, and not to begin their reproduction of the pattern until the entire target exposure cycle was complete, and the fixation cross reappeared. The prepared recording sheet contained numbered groups of 10 open circles which were arranged in horizontal straight lines. After each exposure *O* was required to mark with a pencil in each 10-circle template the circles corresponding to the blackened circles in the pattern. Marking an open circle or failing to mark a filled circle in the pattern both counted as errors in the reproduction of the pattern. The *O*s were allowed to mark responses in any sequence, to change responses before the next stimulus was exposed, and to take as long as they wished to make reproductions of the pattern.

The eight different spatial-temporal conditions of presentation of the patterns are illustrated by a sample pattern in Fig. 1. These conditions were: Right–Left (R–L), in which one element appeared on each exposure of a frame, starting with the first element from the right and proceeding in order to the

---

[2] Manufactured by the Perceptual Development Laboratories, St. Louis, Missouri.

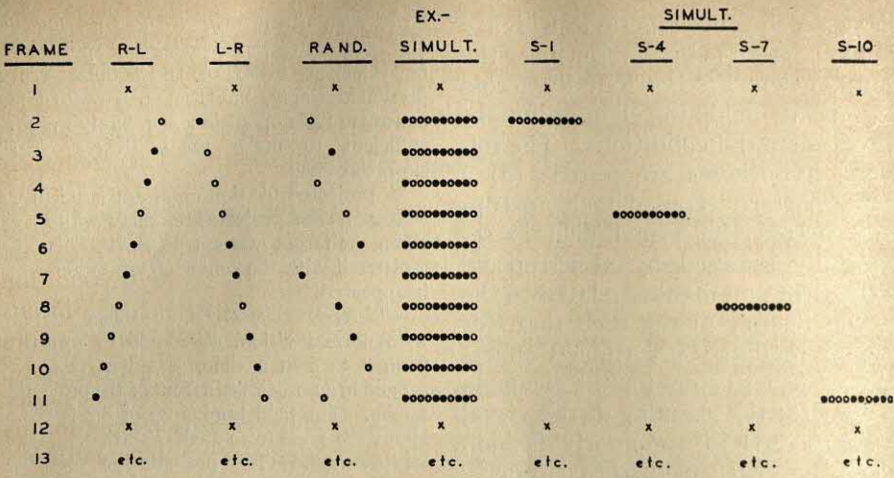


FIG. 1. Illustration of the presentation conditions used in this experiment.

left; Left–Right (L–R), one element per frame, starting with the first element from the left, and proceeding in order to the right; Random (Rand.), one element per frame, starting with any element and proceeding with the remaining elements in a random order; Extended-Simultaneous (Ex.-Simult.), the complete pattern appeared on all 10

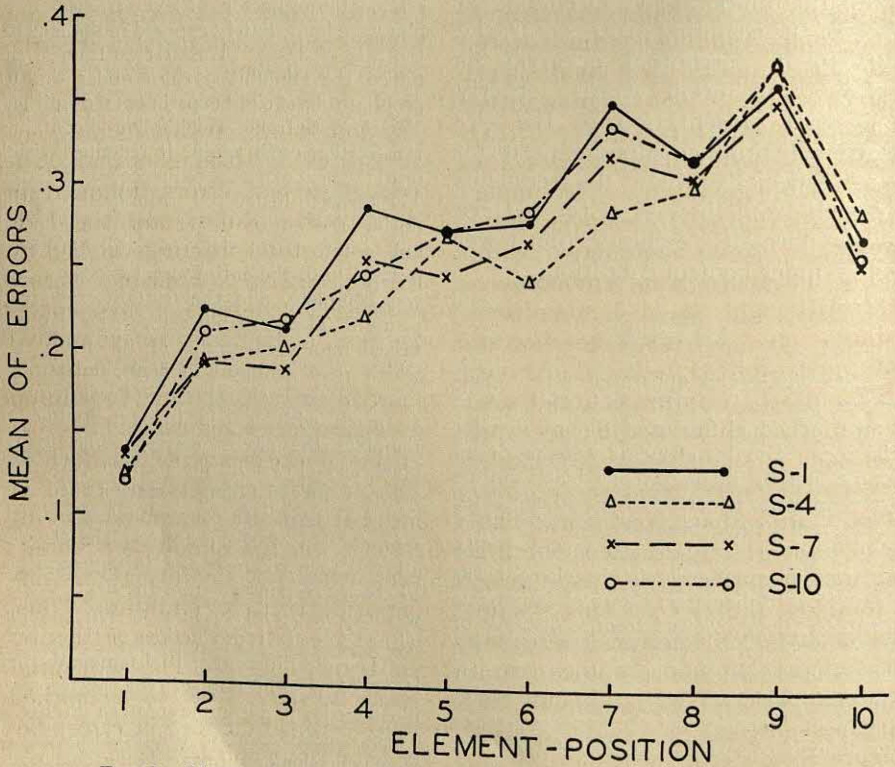


FIG. 2. Means of errors as a function of element position for the four one-frame simultaneous-exposure conditions.

frames; and Simultaneous (Simult.), the complete pattern appeared on only 1 of 4 possible frames in the 10-frame sequence—the first (S-1), fourth (S-4), seventh (S-7), or tenth (S-10). In all conditions, except Ex.-Simult. presentation, the exposure time for each element, therefore, was constant. The four Simult. conditions taken together control for the interval between exposure of the pattern and the reproduction response.

A total of 296 targets was presented to each *O*. This number included 20 targets and their mirror images photographed once for each presentation condition—less the photographs spoiled in the photographing and editing process.

RESULTS

The means of errors per exposure at each element position for each of the Simult. conditions are shown in Fig. 2. Plotted in this manner, these data actually represent proportions of errors, with chance accuracy probably represented by a mean of .5 errors. The various Simult. conditions were by the Friedman two-way analysis of variance (Siegel, 1956) significantly different in overall errors ($\chi_r^2 = 12.73$, $p < .01$). These differences are apparently not produced by a simple decay of perceptual traces over time, however, because the function with varying time delays is not a monotonic one. Although the S-1 condition produces most errors, the fewest errors appear for S-4, followed in order by S-7 and S-10. By inspection, there is no important difference in the shape of the functions; almost every *O* made consistently fewer errors to the left of fixation under each condition. The data combined for the four Simult. conditions show fewer errors to the left for 26 of the 27 *O*s. This result, which is obviously statistically significant, supports previous findings (e.g., Harcum & Dyer, 1962) and confirms Hypothesis *a*.

Figure 3 shows the means for all *O*s of errors per element position under each experimental condition. The data for the Ex.-Simult. condition also substantiates Hypothesis *a*, since 23 of the 27 *O*s exhibit fewer errors for elements in the left field ($p < .0001$, by the sign test). The generally progressive increase in errors for elements from the left end suggests that the scan began from this direction. Results for the Ex.-Simult. condition are similar to those of the Simult. condition, except that more errors appear with only the single exposure. Every *O* showed fewer errors under the Ex.-Simult. condition than for his average under all Simult. conditions. Actually, only 3 *O*s showed fewer errors for even one of the Simult. conditions than for the Ex.-Simult. condition.

Hypothesis *b*—that the scan proceeds in an orderly manner from right to left or from left to right—is not corroborated by data showing the predicted outcome. As can be seen in Fig. 3, overall errors for the R–L, L–R, and Rand. conditions are very nearly equal. Although Cond. R–L produced fewest errors, followed in order by the Rand. and the L–R conditions, this ordering applied to only 6 of the 27 *O*s. Each of the other five possible orderings of these conditions (excluding ties) was represented by at least 3 *O*s. Therefore, contrary to prediction, the Rand. condition did not produce most errors.

Although the group of *O*s as a whole exhibited left superiority—that is, were left-to-right scanners as indicated by the Ex.-Simult. and Simult. curves—the L–R condition was not superior to the R–L condition. Thus, this result is contrary to the prediction from Hypothesis *c*. The sequential exposure results indicate a marked reduction of differences in errors between the right and left halves of the patterns relative to the Simult. conditions. Individual *O*s still exhibited

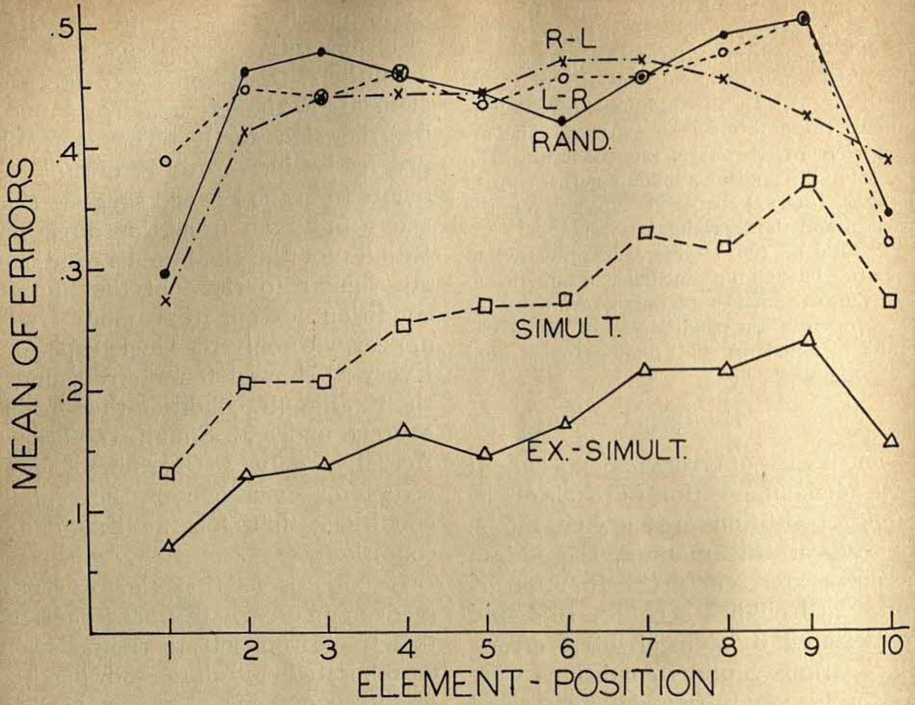


FIG. 3. Means of errors as a function of element position for the five spatial-temporal sequences.

right-to-left differences, however, and a slight nonsignificant superiority for the left half of the patterns in the Rand. condition remains.

An analysis of the data of individual *O*s is required. As is usual with this type of visual task (Ayres & Harcum, 1962; Harcum & Dyer, 1962; Harcum & Rabe, 1958), some *O*s exhibited superior performance for stimuli to the right of fixation, although most *O*s showed greater accuracy for stimuli to the left. Although the present Rand. condition employs sequential presentation, it gives a direct picture (which is uncomplicated by the direction of element presentation) of which side of fixation produces superior perception for an *O*. Therefore, the *O*s may be classified in Table 1 according to whether they show above a median amount of left superiority in the Rand. condition and whether they show above a median amount of relatively greater accuracy for the R–L condition, as compared to the L–R condition. A median test of these data indicates that *O*s who

TABLE 1

RELATIONSHIP BETWEEN THE DEGREE OF LEFT SUPERIORITY IN THE RANDOM CONDITION TO RELATIVE PERFORMANCE WITH R–L AND L–R MOVEMENT OF THE STIMULUS ELEMENTS

	Greater L Superiority	Lesser L Superiority	Total
Greater R–L movement superiority	9	4	13
Lesser R–L movement superiority	4	10	14
Total	13	14	27

demonstrate a trend toward greater accuracy for stimuli in the left hemifield under the Rand. condition also demonstrate a tendency toward greater accuracy under the L–R condition ($\chi_r^2 = 4.46$, $p < .05$). Therefore, the recency effect, which appears for both R–L and for L–R conditions in Fig. 3, is presumably enhanced because the scan begins for some *O*s at the end of the pattern which is presented last. In other words, both presentation-recency and perceptual-primacy favor these elements.

DISCUSSION

The overall superiority of the Simult. conditions to the three conditions involving sequential exposure of the binary elements corroborates the results of Mackworth (1962) with digits. This superiority of simultaneous presentation implies that, if there is a sequential scanning of patterns, the elements of which are all presented simultaneously, the scanning time is very rapid compared to the presentation rate used in these studies. The presentation rate employed by Mackworth was even slower than that used in this experiment. Therefore, in both experiments the immediate visual trace of one element may have decayed to the level of permanent memory before subsequent elements of the pattern were presented. Thus, traces of individual elements could consequently not be easily integrated into a unitary mnemonic pattern for purposes of reproduction. In other words, simultaneous presentation resulted in a composite trace which was scanned while the traces of each element were relatively strong. Sequential presentation, on the other hand, might have resulted in the scanning of one element and nine blank positions on each exposure. This would not foster an integrated perception of traces of individual elements with each trace having nearly equal strength. It might also produce a conflict-of-information, or erasure, process. In either case, an increase in errors with sequential presentation could occur. Also, this might account for the observation of a "recency" effect with sequential left-to-right or right-to-left exposure, rather than the usual "primacy" effect. Thus, we see in Table 1 that the "left-to-right scanners" are better able to profit from the recency of exposure at the left, and the "right-to-left scanners" are better able to profit from the recency of exposure at the right.

The difference between the R–L and the L–R conditions in terms of overall errors is produced primarily by the differences at the first and eighth positions, as can be seen in Fig. 3. For example, if the scan proceeds from left to right on an exposure, but the stimulus sequence proceeds from right to left, and the decay of the trace is rapid relative to the total exposure sequence, then by the time the final element is presented the traces for the preceding elements have decayed to a uniform "permanent memory" level. This level would be somewhat higher for the initial elements of the sequence because of their positional and sequential saliency. In addition, the first elements would be favored by the primacy effect in the scan across the traces.

Relative to Hypothesis *b*, the fact that the Rand. condition did not produce most errors may have been because the stimulus sequence did not consistently conflict with the visual scanning direction for any *O*, whereas a consistent scan in either direction would conflict maximally with either the R–L or the L–R sequence.

Actually, of course, the left superiority in the Simult. and Ex.-Simult. conditions could also be accounted for by the assumptions of a right-to-left scan of the elements plus a recency effect, as Foley's (1957) data would suggest.

The superiority of the Ex.-Simult. condition over the Simult. condition can, of course, be accounted for by the longer exposure duration. This superiority may be greater for the elements of the pattern to the right of the fixation point. One would assume, of course, that the opportunity and likelihood for eye movements to occur during exposure would be greater for the Ex.-Simult. condition.

The results for both of these two control conditions corroborate the results usually found in other studies with simultaneous presentation of this class of stimuli (e.g., Harcum & Dyer, 1962).

The effects of the interaction between stimulus-presentation sequence and *O*'s sequence of marking the elements of the pattern would be difficult to predict. Ayres and Harcum (1962), using simultaneous tachistoscopic presentation of binary patterns such as the ones used in this experiment, have provided evidence for a sequential process in the perception. This process itself produces superior perception on one or the other side of fixation. It also presumably causes the response-marking sequence to proceed in the same direction. The response sequence, in turn, fosters accuracy in the elements reproduced first. On the other hand, in a somewhat different task, Bryden (1960) found that the response accuracy was tied to responding sequence in reproducing a series of forms. Since we did not record the direction with which each *O* reproduced each target, we do not know how responding sequences contributed to the present results. Data from Ayres and Harcum (1962) would suggest that *O*s began marking responses from the end of the pattern which ordinarily for them produced fewer errors in reproduction.

Unfortunately, the question of whether *O*s were able to maintain fixation during the entire exposure sequence in this experiment cannot be certainly answered. However, the presence of decreases in errors for the elements near fixation in the Rand. condition, and in the Ex.-Simult. condition, even though the fixation cross is not reproduced during the exposure of the pattern, would suggest that *O*s did maintain fixation—at least under those conditions.

For reasons cited above, the temporal-spatial characteristics of the perceptual process seem to be confounded in this experiment with mnemonic factors, because the rate of presentation used in this study is apparently slow when compared to the temporal relations in the perceptual process. In short, the lack of correspondence between the sequential-presentation data and the simultaneous-presentation data, either in terms of patterning of errors or totals of errors, implies that the perceptual processes, and possibly other factors, are markedly different in the two situations. This argument appears to us to be a more reasonable way to account for the failure of our predictions in this experiment, than does the hypothesis of a right-to-left, or random, directional scanning of tachistoscopic visual patterns.

REFERENCES

ANDERSON, I. H. The effect of letter-position on range of apprehension scores, with special reference to reading disability. *U. Mich. Sch. Educ. Bull.*, 1946, **18**, 37–40.

AULHORN, O. Die Lesegeschwindigkeit als Function von Buchstaeben und Zeilenlaege. *Pflüg. Arch. ges. Physiol.*, 1948, **250**, 12–25.

AYRES, J. J., & HARCUM, E. R. Directional response-bias in reproducing brief visual patterns. *Percept. mot. Skills*, 1962, **14**, 155–165.

BRYDEN, M. P. Tachistoscopic recognition of non-alphabetical material. *Canad. J. Psychol.*, 1960, **14**, 78–86.

BRYDEN, M. P. The role of post-exposural eye movements in tachistoscopic perception. *Canad. J. Psychol.*, 1961, **15**, 220–225.

CROSLAND, H. R. Superior elementary-school readers contrasted with inferior readers in letter-position, "range of attention," scores. *J. educ. Res.*, 1939, **32**, 410–427.

CROVITZ, H. F., & DAVES, W. Tendencies to eye movement and perceptual accuracy. *J. exp. Psychol.*, 1962, **63**, 495–498.

FOLEY, P. J. The legibility of moving digits as a function of their separation and direction of movement. *Def. Res. Med. Lab. Rep., Toronto*, 1957, No. 76-4.

FRIEDMAN, S. M. Effects of spatial-temporal sequences in visual presentation of binary patterns. *Va. J. Sci.*, 1961, **12**, 202–203. (Abstract)

HARCUM, E. R. Three inferred factors in the visual recognition of binary targets. In J. W. Wulfeck & J. H. Taylor (Eds.), *Form discrimination as related to military problems.* (NAS-NRC Publ. No. 561) Washington, D. C.: National Academy of Sciences—National Research Council, 1957. Pp. 32–37.

HARCUM, E. R. Visual recognition along various meridians of the visual field: II. Nine-element typewritten targets. *U.*

Mich. Engng. Res. Inst. Proj. Mich. Rep., 1958, No. 2144–293-T.

Harcum, E. R., & Dyer, D. W. Monocular and binocular reproduction of binary stimuli appearing right and left of fixation. *Amer. J. Psychol.*, 1962, **75**, 56–65.

Harcum, E. R., & Rabe, A. Visual recognition along various meridians of the visual field: III. Patterns of blackened circles in an eight-circle template. *U. Mich. Engng. Res. Inst. Proj. Mich. Rep.*, 1958, No. 2144-294-T.

Heron, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, **70**, 38–48.

Mackworth, J. F. The effect of display time upon recall of digits. *Canad. J. Psychol.*, 1962, **16**, 48–54.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Terrace, H. S. The effects of retinal locus and attention on the perception of words. *J. exp. Psychol.*, 1959, **58**, 382–385.

Tinker, M. A. Perceptual and oculomotor efficiency in reading material in vertical and horizontal arrangements. *Amer. J. Psychol.*, 1955, **68**, 444–449.

(Received October 6, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 308–310

EXTENDED ACQUISITION TRAINING AND RESISTANCE TO EXTINCTION[1]

SHEPARD SIEGEL AND ALLAN R. WAGNER

Yale University

A group of rats receiving 184 food rewarded, runway acquisition trials extinguished faster than groups receiving only 64 acquisition trials. This was found to be the case whether or not the groups receiving different numbers of acquisition trials were equated prior to extinction with respect to such variables as age, amount of handling, and length of time on deprivation schedule. The results corroborate previous findings by North and Stimmel on the effects of extended acquisition training, and indicate that their results cannot be attributed simply to the age, handling, and deprivation variables.

It is a commonly held position, supported by a variety of studies (e.g., Perin, 1942; Williams, 1938), that resistance to extinction in an instrumental reward situation is an increasing monotonic function of the number of acquisition trials. A recent investigation, however, provides evidence that extended acquisition training can lead to decreased resistance to extinction. North and Stimmel (1960) found that rats given 90 or 135 acquisition trials in a straight runway extinguished faster than rats given 45 acquisition trials. The design of this experiment, however, was such that the various groups started training at the same time, and consequently started extinction at different times. Hence, such variables as age, amount of handling, and length of time on deprivation schedule before extinction were confounded with number of acquisition trials.

The present study was designed to evaluate the importance of these variables in studies of "overlearning," as well as to determine the reproducibility of the effect reported by North and Stimmel.

METHOD

Subjects and apparatus.—The *S*s were 44 experimentally naive male albino rats, 90–110 days old at the start of the experiment.

The apparatus consisted of an enclosed straight runway, 3 in. wide and 4 in. high, divided by guillotine doors into a 12-in. start box, a 30-in. runway, and an 18-in. goal box. Starting and running times were measured, respectively, from the opening of the start door until the breaking of a photocell beam 6 in. distant, and from the breaking of this first beam until the interruption of a second beam 12 in. from the end of the alley.

Procedure.—Prior to experimental training, *S*s received 5 days of habituation, during which time they were reduced to the 75–80% of ad lib. weight at which they were maintained throughout the experiment.

At different intervals following habituation *S*s received acquisition and extinction training in the alley; four trials on Acquisition Days 1 and 2 and eight trials on subsequent days. Daily trials were quasi-distributed by running all *S*s on one trial before starting the first *S* on the next trial. The reward employed during acquisition consisted of four .045-gm. P. J. Noyes pellets.

The *S*s were divided into three experimental groups, differing with respect to (*a*) the number of acquisition trials before extinction, and (*b*) the date of initiation of extinction. Group Long received 184 acquisition trials. The two remaining groups (Group Short-Early and Group Short-Late) received only 64 acquisition trials. Group Long began acquisition training the day following habituation. Group Short-Early also begain acquisition training the day following habituation, and hence began

---

[1] This research was supported by National Science Foundation Grant G-13080.

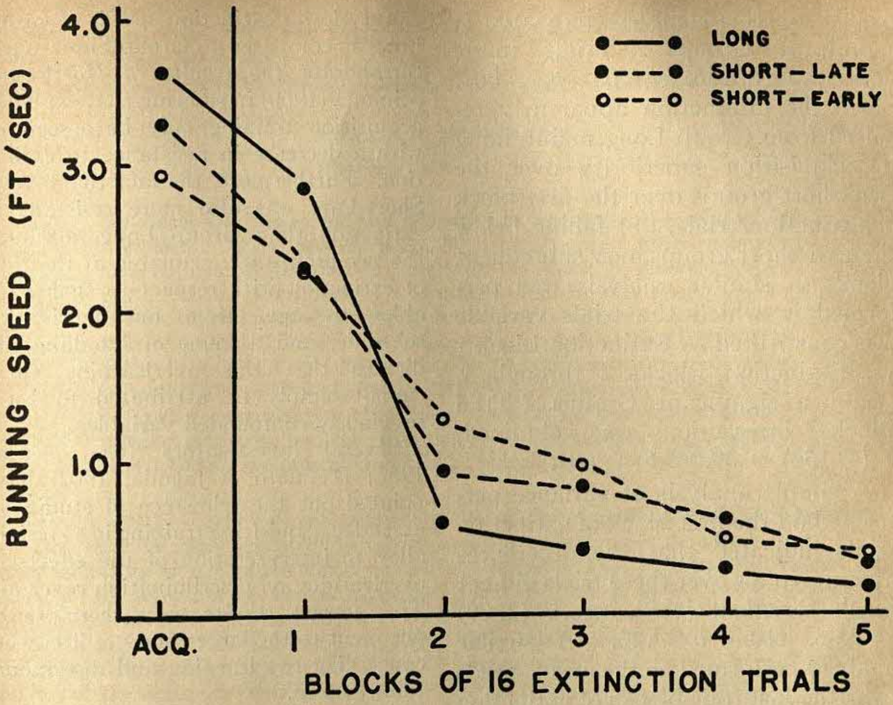


FIG. 1. Mean running speeds for the three experimental groups over the final block of 16 acquisition trials and subsequent blocks of 16 extinction trials.

extinction before Group Long. Groups Long and Short-Early were comparable to groups run by North and Stimmel (1960) in that they differed not only in the number of acquisition trials prior to extinction, but also in age, time on deprivation, and amount of handling. Group Short-Late was not started on runway training until Group Long had completed 116 acquisition trials, enabling the two groups to complete acquisition and start extinction at the same time. Groups Long and Short-Late thus differed in number of acquisition trials prior to extinction, but were equated for age of *S*s and length of time on deprivation. An attempt was also made to equate the amount of handling experienced by *S*s in the latter two groups before extinction. During the time that Group Long was being run, and before Group Short-Late had started, every time an *S* in Group Long was given a trial, an *S* in Group Short-Late was taken out of its cage and placed on a round, elevated platform. After each group had completed acquisition training, it was given 80 extinction trials.

Group Long consisted of 14 *S*s. Fifteen *S*s were assigned to each of the Short-Early and Short-Late groups, but 2 *S*s were discarded from Group Short-Early for failure to learn the approach response.

RESULTS AND DISCUSSION

Figure 1 presents the mean running speeds for the three groups over the final 16 acquisition trials and subsequent blocks of 16 extinction trials. As may be seen, although Group Long ran faster than either of the two Short groups at the end of acquisition and over the first block of extinction trials, it ran slower than either Short group over the last four blocks of extinction.

For purposes of statistical analysis, a mixed design analysis of variance was computed, with experimental groups comprising the between-*S*s factor, and blocks of 16 extinction trials as the within-*S*s factor. The results of this analysis indicated a significant Groups × Trials interaction, $F\ (8, 156) = 6.03$, $p < .001$. Since the interaction apparently resulted from Group Long maintaining its acquisition superiority over the two Short groups over the first block of extinction trials, but falling below the two Short groups over subsequent blocks, a similar analysis was performed in which the trials variable was constituted as Extinction Block 1 vs. Extinction Blocks 2 through 4. Again a significant Groups × Trial Blocks interaction was obtained, $F\ (1, 156) = 39.06$, $p < .001$.

A simple analysis of variance performed on the first block of extinction trials indicated that the differences among groups over these trials did not reach statistical significance, $F\ (2, 39) = 2.36$, $.10 < p < .20$. A similar analysis performed on the mean speeds over the last four blocks of extinction trials, however, indicated a significant difference among the three groups, $F\ (2, 39) = 4.68$, $p < .05$. Subsequent t tests between pairs of groups over these latter trials showed Group Long to be significantly below either Group Short-Early or Group Short-Late, $t\ (25) = 2.54$, $p < .02$ and $t\ (27) = 2.19$, $p < .05$, respectively; whereas the difference between the two Short groups was not reliable ($t = .85$).

Analysis of the starting speed measure revealed the same general trends but a greater variability, and will not be reported here.

The slower extinction speeds of Group Long as compared to Group Short-Early corroborate the results of North and Stimmel (1960) in showing that extended acquisition training may be associated with a decrease in resistance to extinction. Furthermore, the fact that Group Short-Late was also more resistant to extinction than Group Long, although the two groups were equated at the time of extinction with respect to such variables as age, time on deprivation schedule, and amount of handling, indicates that this overlearning effect cannot simply be attributed to these previously confounded variables.

Several investigators (e.g., Amsel, 1962; D'Amato & Jagoda, 1962) have pointed out the relevance of studies of extended acquisition training in a straight alley to interpretation of the effects of overlearning on discrimination reversals. The present results may therefore be pertinent to the larger body of literature concerning overlearning and discrimination reversal.

REFERENCES

AMSEL, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychol. Rev.*, 1962, **69**, 306–328.

D'AMATO, M. R., & JAGODA, H. Overlearning and position reversal. *J. exp. Psychol.*, 1962, **64**, 117–122.

NORTH, A. J., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

WILLIAMS, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–522.

(Received October 17, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 311–313

SUPPLEMENTARY REPORTS

A REPLICATION OF FREE RECALL AND ORDERING OF TRIGRAMS[1]

EDWARD C. CARTERETTE

University of California, Los Angeles

To assess the role of intralist similarity in learning and retention, Ss were shown 12 3-consonant trigrams. One list used 4 different letters (L4, high intralist similarity), the other contained 12 different letters (L12, low intralist similarity). 4 groups of 20 Ss were tested; one in free recall, and another learned only the order of the syllables, for each of the L4 and L12 lists. Conclusions were similar to Horowitz' (1961): (a) free recall for L4 was higher than for L12 early in learning but lower on later trials, and (b) L12 had uniformly higher ordering scores than L4. Horowitz' interpretation that Point a reflects an interaction between intralist similarity and the nature of S's task is weakened by evidence that L12 lists are ordered uniquely by knowing only 1st letters.

Horowitz (1961) sought an explanation for the fact that intralist similarity can have either inhibitory or facilitative effects on the learning and retention of verbal material. To this end he used two lists of 12 three-consonant trigrams, one high similarity list containing only 4 different letters (L4 lists) and one of low similarity containing 12 different letters (L12 lists). Two groups of *S*s recalled as many syllables as they could after each of 10 presentations. Two other groups learned the *order* of the syllables but not the syllables themselves. His most general result was that free recall for L4 was higher than for L12 early in learning (although the reverse was true for the last 5 of the 10 trials), while L12 was consistently easier to order than L4. The ordering experiment showed L12 superior to L4 over all 10 trials. Horowitz interpreted his results to indicate an interaction between intralist similarity and the nature of *S*'s task. This interaction may be important in understanding the method of serial anticipation with which recall and intralist do not appear to be systematically related.

This replication sprang from a need for comparative data (Horowitz' *S*s were grammar school children) prior to a series of experiments on the role of sequential redundancy in the language of children. Horowitz used a single intact class for each of the four conditions; it was desirable to avoid the danger of selection and correlation inherent in the use of groups. Finally, it was suspected that ordering of the L12 lists was especially easy because knowledge of only a syllable list's first letters was enough to order the set.

Method.—The consonant trigrams used were identical to those of Horowitz (1961).

Each of the four list-condition combinations was presented in a random order to groups whose size varied from 2 to 5 *S*s. Thus a given group was run under one of the four conditions: free recall, L4 or L12; ordering, L4 or L12. There were 20 *S*s of both sexes in each of the four groups, who were volunteers from classes in introductory psychology at the University of California, Los Angeles.

The introductory instructions, the same for all groups, were read aloud by *E*.

> I am going to show you a number of cards with a different set of letters printed on each one. I will show you one at a time until I have finished the set. After I have shown you the complete set of cards, you will be asked some questions about what you have seen, so pay close attention to each card as it comes up. We will go through the set of cards several times.

After *S*s had seen the entire list once, specific instructions were given to the two groups. Each trigram was shown for 4 sec.

> [Free recall] Will you please write down on the card I pass out as many of the syllables as you can remember in any order. Feel free to guess. Take your time and try to get as many as possible. When you are through turn your card over and I will take them all up at once.
>
> [Ordering] I am now going to give each of you a small set of cards. On each card is

---

[1] Grateful acknowledgment is made to Jane Ebersole for collecting and organizing the data.
This work was supported by a grant (MH-06986-01) from the National Institutes of Health.

TABLE 1

Mean Correct Recall Scores and Mean Correct Placement Scores with L4 and L12

Trial	Mean Syllables Correctly Recalled		Mean Correct Placements	
	L4	L12	L4	L12
1	6.35	1.65	1.60	1.80
2	5.80	3.75	2.60	4.90
3	5.40	4.80	3.00	7.50
4	5.65	5.35	3.30	9.60
5	5.95	6.40	4.20	9.80
6	5.50	7.50	4.10	10.70
7	6.75	7.85	5.30	11.10
8	6.65	8.20	5.30	11.50
9	6.70	8.85	6.30	11.90
10	6.75	8.75	6.00	11.90

> printed one of the syllables you have just seen. Your job is to put these cards in the same order as those you just saw. Feel free to guess. Take your time and try to be as accurate as possible. When you are through, set the pack of cards in front of the desk and I will pick it up. We will then go through the cards again.

Horowitz did not say what instructions were given to his *S*s.

Results and discussion.—The most general result is identical with that of Horowitz, namely (*a*) free recall for L4 is higher than for L12 early in learning and this superiority is lost on later trials (but see the *caveat* below)

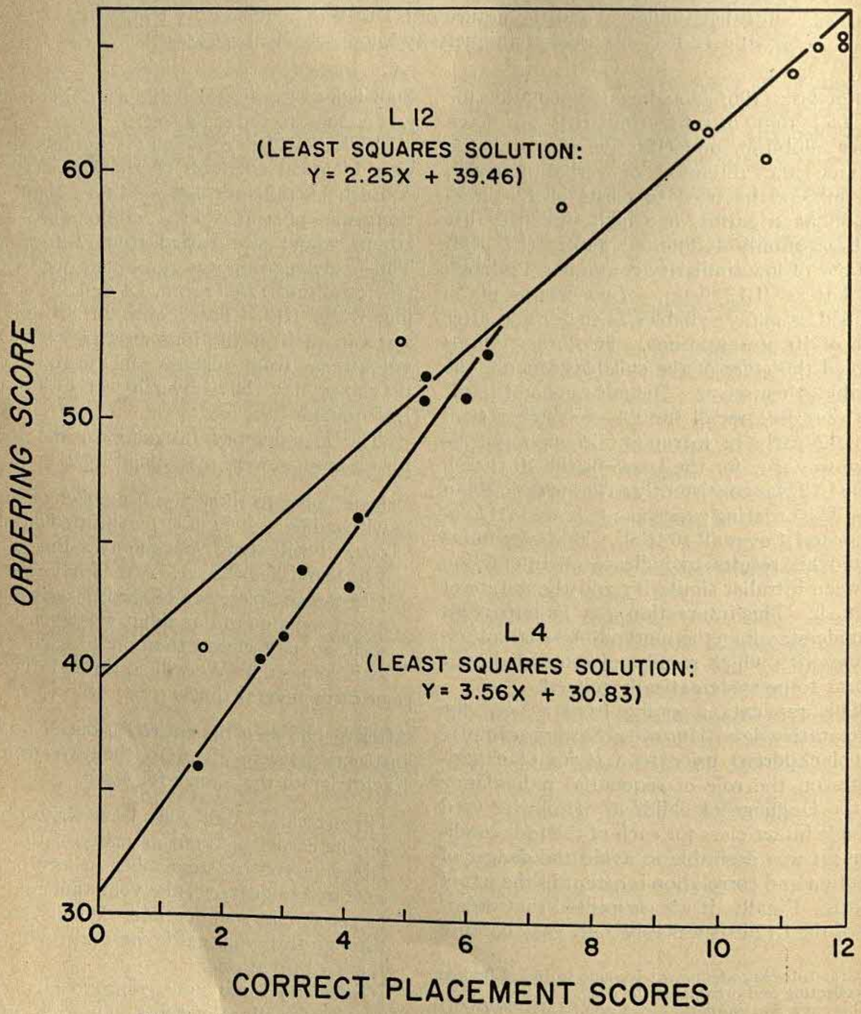


FIG. 1. Correct placement scores plotted against mean ordering scores.

and (*b*) L12 has uniformly higher ordering scores than L4.

Table 1 gives the mean number of syllables correctly recalled in each trial. L4 is superior over the first four trials. This initial difference is significant at $p < .001$ by a Mann-Whitney *U* test. Note, however, that there is no improvement from Trial 1 to Trial 10. A statistical test rejected the hypothesis that deviations from a straight line were of quadratic or higher order and that the slope of the function was nonzero, since the computed slope was well within the 95% confidence interval of zero. Horowitz' L4 rises smoothly from a mean of 3.12 to 7.8 over the 10 trials. This is the major difference between his and the present results. It is not possible to decide between the two most likely alternative causes, differences in *S*s or instructions. However, guessing was explicitly encouraged in the present study, and the mean correct recall score on the first trial is consistent with the assumption of guessing.

The analysis of variance of correct recall scores showed that the mean difference between L4 and L12 lists was not statistically significant at the .05 level, $F\,(1,\ 38) < 1$, even though the Lists × Trials interaction was significant, $F\,(9,\ 342) = 19.84$, $p < .001$. The effect of trials was significant as would be expected, $F\,(9,\ 342) = 13.20$, $p < .001$.

Turning to the ordering data, it is easy to see (Table 1, third and fourth columns) that the mean placement scores for L12 were uniformly superior to those for L4. There is remarkably good agreement between the ordering data in the two experiments. The plot of correct placement scores against mean ordering scores. (Kendall's *P*, 1955, pp. 3–7) from the present experiment is given in Fig. 1, together with the least squares solution. Horowitz' solution for his L12 data was $Y = 1.91X + 44.03$, and for the L4 data, $Y = 3.46 + 31.99$. These differ very little from ours in either slopes or intercepts.

On the basis of serial position curves, Horowitz argues that "the highly similar items of L4 . . . are not ordered relative to one another, but are memorized individually," for if the first, second, and last items were memorized and the remaining nine ordered at random, the mean correct placement score would be 4.0. His L4 and L12 *S*s give mean scores of 1.44 and 4.7 on Trial 1 and thus he infers a different process to account for L4. However, if one assumes that the position of the *first letter* in each syllable is learned for the first, second, and twelfth items, and that the second syllable is guessed randomly, and the remaining nine items are ordered at random, the expectation is $(\frac{1}{3})(3 + 1) = 2$ for L4 and $(3 + 1) = 4$ for L12, not too different from his obtained values. All but one *S* in the present study stated that by the end of the experiment ordering was done on the basis of the first letter alone. The further argument of Horowitz that a "true" serial position effect does not arise for the L4 lists until Trial 8 seems plainly contradicted by the Trial 5 plots for L4 and L12 of serial position vs. correct placements (Horowitz, 1961, p. 55, Fig. 1).

In general these findings are in agreement with Horowitz' in spite of differences in *S*s and method. A stronger conclusion than his seems reasonable, namely that L4 lists (high intralist similarity) are inferior to the L21 lists (low intralist similarity) in both free recall and ordering. The view that the apparent initial superiority in free recall of L4 and L12 may be due to guessing is strengthened by these data obtained under instructions favorable to guessing. The strong linear relationship between ordering and correct placement scores, the product-moment correlations being .976 for the L4 and .972 for the L12 groups, together with the very close agreement between these and Horowitz' findings suggests that the effort in computing the ordering score is excessive.

REFERENCES

HOROWITZ, L. M. Free recall and ordering of trigrams. *J. exp. Psychol.*, 1961, **62**, 51–57.

KENDALL, M. G. *Rank correlation methods.* London: Griffin, 1955.

(Received September 20, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 314–316

EFFECTS OF GROUPING OF STIMULI ON THE PERCEPTION OF RELATIVE FREQUENCY[1]

DWIGHT E. ERLICK

Aerospace Medical Laboratory, Wright-Patterson Air Force Base, Ohio

The ability to perceive the overall relative frequency of 2 groups of random sequential categories as a function of the number of categories per group was investigated. The number of categories per group was the same for the 2 groups being compared, i.e., either 1 (C_1), 2 (C_2), or 4 (C_4). Letter sequences were used to represent categories within a group. The S's task was to determine whether the combined total frequency of A or C was greater (C_1); or D + E or G + H was greater (C_2); or Q + R + S + T or W + X + Y + Z was greater (C_4). All the letters comprising both groups being compared were presented in a random sequence 1 at a time at the rate of 4 per sec. The results indicate an approximately linear function between the information transmitted (T) and the frequency difference between the groups of categories (ΔF) for each of the 3 C conditions. The slope $T/\Delta F$ decreased in a linear fashion as a function of C; $\text{slope}_{T.\Delta F} = -.0020\,C + .0133$.

Previous investigations have been concerned with comparing the relative frequencies of *individual* categories (Erlick, 1961, 1962). This investigation was designed to measure the perception of the overall relative frequency of two groups of categories as a function of the number of different categories per group.

Method.—The number of different categories (C) comprising a group was either one (C_1), two (C_2), or four (C_4). Letters of the alphabet were used to represent categories. To equate the association value of categories within a group, letter sequences of the alphabet were used. Relative frequency was based on comparison of the frequencies of two groups of categories with the same number of categories being used for both groups being compared within each C condition. For each of the C conditions, *S*'s task was to designate the group having the greater total frequency. Letters used for the C conditions were as follows: C_1, A-C; C_2, D, E vs. G, H; and C_4, Q, R, S, T vs. W, X, Y, Z. The combined frequency of the categories of the less frequent group (G_L) was always 20, and the categories comprising the more frequent group (G_M) had an increased frequency at one of four values. The percentage increments in frequency (ΔF) between G_L and G_M are designated by

$$\Delta F = \frac{G_M - G_L}{G_L} \times 100.$$

As C increased the ΔFs were increased and covered a range from 10 to 120. For C_2 and C_4, the frequencies of the different letters comprising a group were determined by four different random selections, with the stipulation that a minimum of one of each of the different letters must occur. Eight different random sequences of all the letters in G_L and G_M were presented for each ΔF condition within each C condition. The G_L and G_M conditions for each ΔF within each C condition were counterbalanced in terms of the letter symbols used.

The letters were presented one at a time with equal on-off times at the rate of four letters per second. The letters comprising the various groups for each C condition were posted below the viewing screen, to facilitate *S*'s memory as to which letters belonged to the same group.

A trial required one judgment and consisted of observing a random series of either two, four, or eight different letters, with the appropriate combinations of half the letters adding up to 20 and the remaining half having one of the ΔF increments in frequency. A session consisted of 16 trials of the same C condition, counterbalanced so that each of the four ΔFs preceded and followed every other ΔF, including itself, once. An experimental period, which lasted about 1½ hr., including rests between sessions, consisted of six sessions (two presentations of each C condition) presented according to a replicated Latin square counterbalancing order, i.e., ABCCBA, CABBAC, and BCAACB.

---

[1] The data were collected at the Engineering Psychology Research Project, Antioch College.

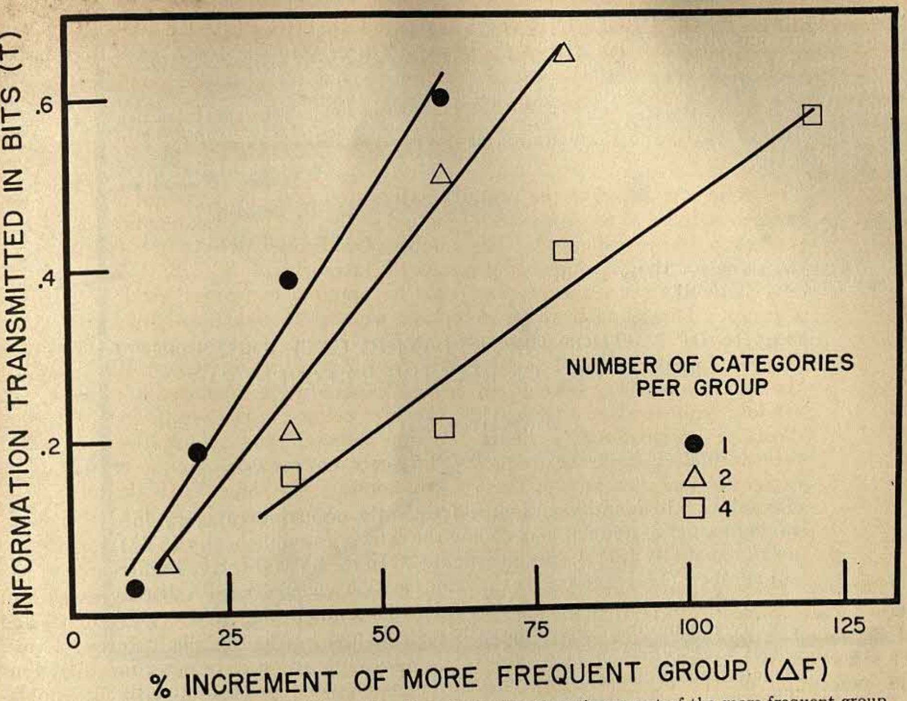


FIG. 1. Information transmitted (T) as a function of the percentage increment of the more frequent group (ΔF) for groups having one, two, and four categories. (Each point is based on 592 observations.)

The *S* was instructed to write the combination of letters of the group that appeared more frequently at the end of each trial, i.e., in the C_1 condition either A or C; for C_2 either DE or GH; and, for C_4 the first and last letter of the sequence of four letters, i.e., QT or WZ. Each *S* made a total of 32 judgments at each of the three C conditions.

Results and discussion.—The following analyses are based on the total responses of all 74 *S*s. The percentage of correct identifications for each of the four ΔF values at each of the three C conditions was converted into bits of information transmitted (T). This measure assumes that both groups, i.e., G_L and G_M, within a C condition have equal likelihood of being selected as G_M, resulting in an information input of 1 bit (Erlick, 1962).

An examination of the relationship between information transmitted (T) and the frequency difference (ΔF) between groups for each of the three C conditions gave the following equations based on the method of least squares (see Fig. 1). The T subscript indicates the number of categories (C) comprising a group.

$$T_1 = .0112\Delta F - .0479$$
$$T_2 = .0095\Delta F - .0922$$
$$T_4 = .0052\Delta F - .0436$$

The slope T/ΔF systematically decreases as C increases. However, the intercepts show no systematic variation and are approximately equal. An examination of the relationship between the slopes as a function of C indicates a linear relationship and gives the following equation based on the method of least squares:

$$\text{Slope}_{T.\Delta F} = -\ .0020\ C + .0133$$

where C indicates the number of different categories per group. Combining this equation with the slope of the general equation and intercept for an information input of 1 bit ($T = .0113\Delta F - .0726$; Erlick, 1962), gives the following equation relating information transmitted (T), frequency difference between two groups of categories (ΔF), and the number of categories per group (C), for one, two, and four categories per group: $T = (-\ .0020C + .0133)\ \Delta F - .0726$. The standard error based on predicting T for all three C conditions: $\sigma_{T.\Delta F} = .0371$.

The fact that C_1, C_2, and C_4 have approximately linear relationships indicates that the same functional relationship holds when the relative frequency differences (ΔF) are based on groups of categories, at least when dealing with two groups composed of either two or four categories. An analysis of the three C equations allows the conclusion that information transmitted decreases in direct proportion to an increase in the number of different categories per group. The small standard error of estimates for the linear equation relating T, ΔF, and C substantiates this relationship.

REFERENCES

ERLICK, D. E. Judgments of the relative frequency of a sequential series of two events. *J. exp. Psychol.*, 1961, 62, 105–112.

ERLICK, D. E. Perception of the most frequent category of a random series as a function of the number of categories. *J. exp. Psychol.*, 1962, 63, 115–118.

(Received October 23, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 3, 316–318

EFFECTS OF METHOD OF DISPLAYING CATEGORIES ON THE PERCEPTION OF RELATIVE FREQUENCY[1]

DWIGHT E. ERLICK

Aerospace Medical Laboratory, Wright-Patterson Air Force Base, Ohio

The ability to identify the more frequently occurring category in a random series as a function of coding the categories symbolically (SYM) and/or spatially (SPA), was investigated. In the SYM display, 4 different numbers (1–4) appeared in the same location while in the SPA display the symbol O appeared in 4 different locations, and in the SYM + SPA display the numbers 1–4 appeared in 4 different locations. The categories were presented randomly at the rate of 4 per sec. Ss were instructed to indicate the category which appeared most frequently. The results show significantly better performance for the SPA and SYM + SPA displays over the SYM display and no significant difference between the SPA and the SYM + SPA displays. The results are explained in terms of the facilitation of handling externalized physical representations of concepts as opposed to abstract representations.

In previous investigations, categories have been represented, i.e., coded, symbolically as letters or numbers (Erlick, 1962). Also, all categories being compared as to relative frequency were presented on the same spatial location of a 3 × 4 in. frosted plastic screen. Thus, the only cue *S* had to discriminate which category had occurred was the symbolic representation of that category.

Investigations (Bradley, 1959; Eriksen & Hake, 1955) have shown that on some tasks there is an increase in accuracy in discrimination, when multiple cues are available to identify the stimuli. The purpose of this experiment was to determine the ability to perceive the relative frequency of random categories as a function of the categories being coded: in a symbolic dimension; in a spatial dimension; and, in both dimensions at the same time. From previous evidence it would be hypothesized that the dimension which was more discriminable would also give greater accuracy in terms of identifying the correct category. Therefore, it would be hypothesized that the spatial display would be easier to categorize since it has a distinct physical representation of each category. This might be expected to give a higher accuracy of identification as opposed to the symbolic representation which requires the sorting process to be more "internalized."

On the other hand, the addition of a symbolic representation of a category to an already physically separated spatial representation would not be expected to add any additional discriminatory cues. Therefore, it would be hypothesized that there would be no difference in accuracy between the spatial display and the spatial plus symbolic display.

Method.—The *S*'s task was to determine which of four random sequential symbols (categories) had occurred most frequently. Three of the categories always had the same frequency, 20, while the fourth had a frequency of one of six values; 24, 26, 30, 34, 38, or 44. The ΔFs ranged between 20 and 120 (see Erlick, 1962).

Three different displays were used to

---

[1] The data were collected at the Engineering Psychology Research Project, Antioch College.

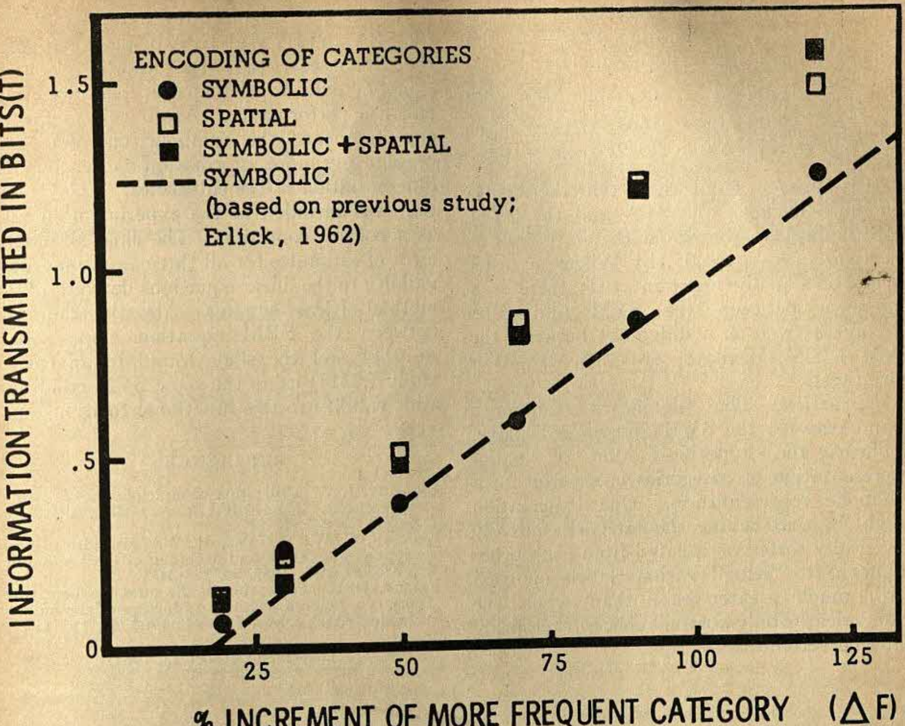


FIG. 1. Information transmitted (T) as a function of the percentage increment (ΔF) of the more frequent category for symbolic, spatial, and symbolic plus spatial coded categories. (Each point is based on 300 observations.)

present the four categories. In the symbolic display (SYM), the four categories were represented by the numerals 1, 2, 3, and 4 and all four numbers appeared on the same spatial location of one 3 × 4 in. frosted plastic screen. In the spatial display (SPA), each of the four categories was represented by the same symbol "O" which appeared on four *different* 3 × 4 in. frosted screens.

The symbolic plus spatial display (SYM + SPA) was the same as the SPA display, with the exception that instead of the same symbol "O" appearing on each of the four different display screens, the numerals 1, 2, 3, and 4 each appeared on a different screen, the same number always appearing on the same screen, i.e., 1 on the leftmost screen progressing to 4 on the rightmost screen. Thus, the SYM display consisted of one 3 × 4 in. screen while the SPA and SYM + SPA displays consisted of four 3 × 4 in. screens arranged horizontally into one 12 × 4 in. display. For all display procedures only one symbol appeared at a time. Each of the three different displays was presented to a different group of 50 *S*s. A trial required one judgment and consisted of observing a random series of the four categories with one of the six ΔF increments. Each group was presented 36 trials, counterbalanced in terms of the sequences of the different ΔFs. For each of the different display dimensions, the four different categories were counterbalanced so that each occurred as the more frequent (C_M) category the same number of times.

Results.—The following analysis is based on the summated responses of each of the three groups of 50 *S*s. The percentage of correct identifications for each of the six frequency increments (ΔF), for each of the three display dimensions were converted into information transmitted (T) (Erlick, 1962). A graphic analysis based on this measure indicated no systematic interactions between preferences for symbols or spatial locations and the ΔF conditions. Therefore, the responses to the counterbalanced C_M for each ΔF within each display dimension were combined in the following analysis.

An examination of the relationship between information transmitted (T) and the frequency difference (ΔF) between categories for the SYM, SPA, and SYM + SPA displays gave the following equations based on the method of least squares (see Fig. 1).

$$T_{SYM} = .0114\Delta F - .1572;\ \sigma_{T.\Delta F} = .0380$$
$$T_{SPA} = .0144\Delta F - .1668;\ \sigma_{T.\Delta F} = .0571$$
$$T_{SYM+SPA} = .0153\Delta F - .2379;\ \sigma_{T.\Delta F} = .0525$$

It is apparent from Fig. 1 that there is no difference between the SPA and the SYM + SPA displays (based on the six points of the two curves $p > .05$, by Wilcoxon test.) It is also equally apparent that there is a difference between the SYM and SPA displays as well as a difference between the SYM + SPA displays ($p < .05$, by Wilcoxon test).

Discussion.—The significant difference found between the SYM and SPA displays, confirms the hypothesis that a spatial representation of categories is superior to a symbolic representation. One explanation might be that having the categories already physically sorted or isolated from each other allows *S* to "hang" various labels on each with much greater ease than when the categories are only abstract concepts requiring internal verbalizations.

The lack of a significant difference between the SPA and SYM + SPA displays supports the hypothesis that the addition of the symbolic cue to the spatial cue does not enhance performance. Again this might be expected since the symbolic cue does not present any more information concerning the differentiation of the categories. This condition was included in the experiment mainly as a control condition. The small standard error of estimates for all three equations adds validity to the linear equations derived by the method of least squares. Also the similarity between the SYM equation slope $T/\Delta F = .0114$ and the slope found in an earlier study (.0113) using the same SYM condition adds validity to this functional relationship.

REFERENCES

BRADLEY, J. V. Utilization of multiple cues in paired comparisons. *USAF WADC tech Rep.*, 1959, No. 59-548.

ERIKSEN, C. W., & HAKE, H. W. Multidimensional stimulus differences and accuracy of discrimination. *J. exp. Psychol.*, 1955, 50, 153-160.

ERLICK, D. E. Perception of the most frequent category of a random series as a function of the number of categories. *J. exp. Psychol.*, 1962, 63, 115-118.

(Received October 23, 1962)

Journal of
Experimental Psychology

VOL. 66, NO. 4 OCTOBER 1963

STIMULUS INFORMATION AND CONTEXTUAL INFORMATION AS DETERMINANTS OF TACHISTOSCOPIC RECOGNITION OF WORDS

ENDEL TULVING AND CECILLE GOLD[1]

University of Toronto

Tachistoscopic word recognition can be conceptualized as a situation in which S has to select a previously learned response from a set of alternative responses. Information from various sources can be used by S in arriving at the correct response. 3 experiments, in which recognition thresholds for target words were correlated with measures of information content and degree of congruity of their pre-exposure contexts, were conducted to evaluate the hypothesis that the amount of information that S needs from the tachistoscopically presented stimulus varies inversely with the amount of available relevant contextual information. Data from the experiments were interpreted as providing support for the hypothesis.

While it is perhaps rash to suggest that the human organism is nothing but a communication channel, conceptual analyses of behavior are often aided by analogies between experimental paradigms and communication situations. Consider, for instance, the experiment in which recognition thresholds are determined for tachistoscopically presented words. The *S*'s task in the experiment can be conceived of as the selection of responses from a more or less well-defined set of previously learned and integrated responses. This response selection is guided by various sources of information, both controlled and uncontrolled, that are available to *S*.

One such source of information lies in the tachistoscopically presented stimulus word. Probability of correct response increases monotonically with exposure duration, expressing the direct relation between amount of stimulus information available from the tachistoscopic exposure and the duration of the exposure, but the amount of information available is relatively independent of the amount of information contained in the stimulus (Miller, Bruner, & Postman, 1954; Tulving, 1963).

Given a systematic relation between exposure duration and the amount of information available from the stimulus, measures of recognition threshold can be thought of as reflecting the amount of stimulus information that *S* needs for response selection. In these terms, then, all systematic variability in recognition thresholds indicates differential availability of information from other sources.

---

[1] Now at Goucher College.

The *S* can sometimes select the correct response in the complete absence of stimulus information (Goldiamond & Hawkins, 1958; Smock & Kanfer, 1961). This observation, and the fact that complete but incorrect responses conveying some information often occur prior to the veridical report (Bricker & Chapanis, 1953), provide ample evidence for the utilization of information from other sources. The existence of these sources has long been known, and terms such as set, hypothesis, and expectancy have been used to refer to them.

It is reasonable to assume that different sources of information are complementary to one another in the sense that if one source provides much information then less information is needed from other sources. The experiments reported in the present paper were designed to evaluate the hypothesis that visual duration threshold (VDT), reflecting the amount of stimulus information needed by *S*, varies inversely with the amount of information provided by the context of the target word. In the first experiment, VDT was measured as a function of length of preexposure contexts that were either congruous or incongruous with the target word. The second experiment provided quantitative measures of two properties of contexts: information content, and degree of congruity with the target word. In the third experiment, covariations among VDT of target words and the two dimensions of contexts were examined.

EXPERIMENT I

Grammatical sequences of words and letters are redundant: *S* can often guess the word or letter that follows a given sequence. The accuracy of such guesses increases with increasing length of preceding context, up to a limit of 5–10 words (Aborn, Rubenstein, & Sterling, 1959) or 30–35 letters (Burton & Licklider, 1955). Increasing the length of context narrows the range of possible alternatives or increases the amount of information that is available to *S* about the target word. If, in a tachistoscopic experiment, VDT is directly related to the amount of stimulus information needed by *S* to make the correct response, and if the amount of this information varies inversely with the amount of relevant information available from other sources, then it would be expected that VDT for words appearing in presence of context would decrease with increasing length of the context.

Increasing the length of context increases the amount of relevant information about the target word only if the context is congruous with the target word, that is, if the word following a given context does not violate the rules of the language with which *S*s are familiar. If the target word is incongruous with the context, then increasing the length of the context would be expected to decrease the amount of relevant information and to increase recognition threshold for the target word shown tachistoscopically.

Method

General design.—Thresholds for target words were determined under 10 experimental conditions provided by orthogonal combinations of five lengths (0, 1, 2, 4, and 8 words) and two levels of congruity ("congruous" and "incongruous") of context. Each of the 10 *S*s employed in the experiment was tested

TABLE 1

CONTEXTS AND CORRESPONDING TARGET WORDS

Sentence No.	Context	Target Word
1	The actress received praise for being an outstanding	PERFORMER
2	Three people were killed in a terrible highway	COLLISION
3	The escaped soldier was captured and court-martialed for	DESERTION
4	Far too many people today confuse communism with	SOCIALISM
5	She likes red fruit jams of strawberry and	RASPBERRY
6	The skiers were buried alive by the sudden	AVALANCHE
7	Many colorful flowers and stately elms lined the	BOULEVARD
8	Medieval knights in battle were noted for their	GALLANTRY
9	More money buys fewer products during times of	INFLATION
10	The loud piercing screams occurred with regularly increasing	FREQUENCY

under all conditions. Ten different target words, selected on the basis of a small pilot study, were used throughout the experiment.

Stimulus words and contexts.—Twenty-nine nine-letter, three-syllable nouns occurring four times per million (Thorndike & Lorge, 1944) were selected for pretesting in order to arrive at a relatively homogeneous group of target words. For all these words VDTs were obtained from 11 *S*s, graduate students in psychology. The order of presentation of the words was systematically changed over *S*s. On the basis of the data from this small pilot study, 10 words, having similar median thresholds and showing least inter-*S* variability, were selected as target words for the main experiment.

Each of the 10 target words was used in construction of a 9-word sentence, with the target word as the final word in the sentence. The preceding words in the sentence constituted the 8-word congruous context for the target word. These contexts and their corresponding target words are shown in Table 1.

Contexts of Lengths 4, 2, and 1 were constructed by omitting the first four, first six, or first seven words, respectively, from each eight-word context. Absence of any context constituted the condition of Length 0.

Incongruous contexts of various lengths for a given target word consisted of congruous contexts of other target words. This procedure was based on the assumption that all target words, except the one that was used in the construction of a given sentence, would be incongruous with other contexts.

Subjects and procedure.—Ten female undergraduate psychology students served as *S*s in this experiment. Their age range was from 18 to 23 yr., with the median age of 20. None had had any experience in psychological experiments. They had normal vision, corrected or uncorrected.

Each *S* was tested individually. Instructions to *S*s included a statement that the words they were to identify "actually follow sequences of words," that parts of these sequences would be shown to them on most trials, and that the knowledge of the sequences may help them to recognize the words. All *S*s were encouraged to guess when in doubt about the identity of the target word. VDTs for four practice words were determined for each *S* prior to the main part of the experiment in order to absorb the most striking practice effects.

The contexts were typed in upper case letters and shown to *S*s in the upper part of the pre-exposure field of the Gerbrands mirror tachistoscope. The target words were also typed in upper case letters and presented for controlled durations in the exposure field. The level of background luminance in both fields was held constant at 2 ft-c (Macbeth illuminometer reading).

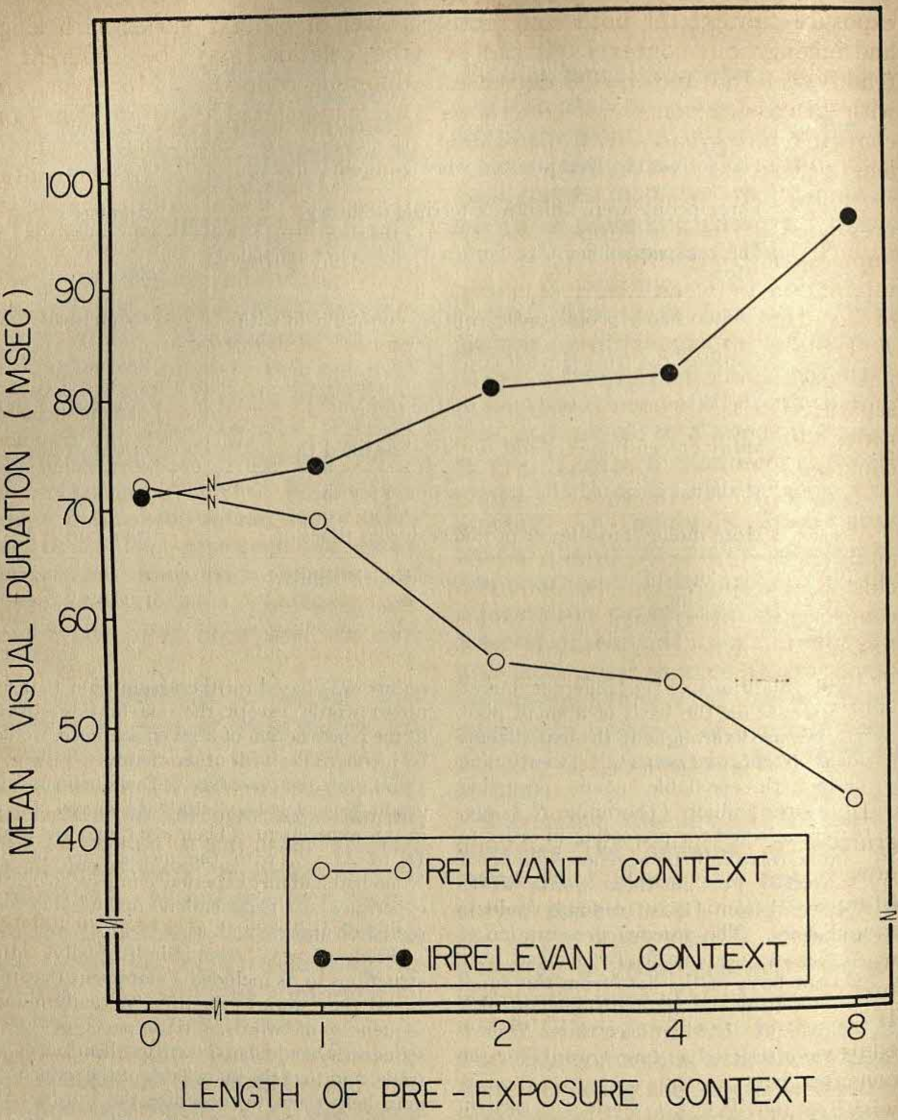


FIG. 1. Visual duration threshold as a function of length of context for congruous and incongruous contexts.

VDTs were measured by the ascending method of limits. Each target word was exposed initially at 10 msec. and exposure durations were increased in successive steps of 10 msec. until *S* gave the veridical report. At each level of duration two exposures were given. If *S* correctly identified the word on the first exposure at 60 msec., for example, VDT was recorded as 60 msec., if identification followed the second exposure at that level, VDT was recorded as 65 msec.

The order of presentation of different lengths of context of target words was systematically counterbalanced over *S*s. The order of presentation of congruous (C) and incongruous (I) contexts, however, followed the same arbitrarily selected fixed pattern for all *S*s: CICCIICICI.

Results

The results of Exp. I are summarized in Fig. 1. Mean VDT is plotted as a function of length of pre-exposure context for both congruous and incongruous contexts. It can be readily seen that mean VDT decreases with increasing length of congruous context and increases with increasing length of incongruous context.

Analysis of variance of VDT measures adds statistical support for the impression gained from Fig. 1. The interaction between length and type of context is highly significant: $F(4, 81) = 12.35$, $p < .01$. Individual analyses of the two curves separately also yield significant F ratios for length of context: for congruous context, $F(4, 35) = 11.47$, $p < .01$; for incongruous context, $F(4, 35) = 4.44$, $p < .01$. These analyses permit us to draw the conclusion that increasing the length of congruous context facilitates tachistoscopic identification of words, while increasing the length of incongruous context interferes with it.

EXPERIMENT II

In Exp. I only two levels of congruity were examined. It is obvious, however, that congruity is a continuous variable rather than a dichotomous one. Furthermore, the data were averaged over 10 different contexts at each level of length of context. It is quite possible that different contexts of a given length vary considerably in degree of congruity with respect to selected target words, both within the set of congruous and within the set of incongruous contexts. Quantification of degree of congruity between each context and its designated target word would presumably permit a somewhat more detailed analysis of the relation between congruity and VDT.

Similar considerations are relevant to the second dimension of context varied in Exp. I, namely its length. While it is clear that the constraining power of context varies with length, the relation may be different for different contexts. Moreover, since we manipulated length in Exp. I only in order to vary the amount of information conveyed by the context, it seemed useful to determine the constraining power of various individual contexts.

Measurement of Congruity and Contextual Information

Degree of congruity between a context and a given target word can be measured simply in terms of relative frequency with which the target word occurs as a response in a large group of *S*s who receive no stimulus information, that is, in a guessing situation. If all *S*s guess the target word correctly, the congruity is maximum, if none of them does so, the congruity is minimum. If each *S* is permitted only one guess, then the measure of congruity (C) is given by the proportion of *S*s who guess the target word correctly.

Amount of the constraining power of the context, that is the amount of information conveyed by the context, can vary independently of congruity, since it is specified without regard to the target word. It can be estimated from the distribution of responses in different response categories given by a large group of *S*s. Many different measures could be used to assign a numerical value to the information content (I) of a context. In the present experiment, the following measure was used:

$$I = \frac{\Sigma n_i \log n_i}{N \log N} \qquad [1]$$

In this formula, n_i refers to the frequency of a given response i in the distribution of N responses, where each of N individuals gives one response. The formula is a restatement of the formula for redundancy in a set of events as used in the mathematical theory of communication (McGill & Quastler, 1955):

$$I = 1 - \frac{H(x)}{\max H(x)} \qquad [2]$$

The measure of *I*, as given by Formula 1 can vary between zero and unity, the former value representing the maximum amount of uncertainty in *S*s' responses and hence minimum amount of information in the context, and the latter value representing minimum uncertainty in *S*s' responses and hence maximum amount of information conveyed by the context.

Method

Each of the 50 contexts used in Exp. I (5 lengths × 10 sentences), followed by a blank space, was mimeographed on a small slip of paper. Different slips were stapled in haphazardly varied orders into small booklets. Each booklet contained 10 slips. Five of these were irrelevant for the purpose of the present experiment, although similar in nature. The other 5 represented five different lengths of context associated with different stimulus words. Ten booklets exhausted all 50 contexts used in Exp. I. A total of 1,000 booklets was prepared.

The *S*s were 1,000 undergraduate students enrolled in various psychology courses at the University of Toronto. Testing was conducted in six large classes as a part of regular classroom procedure. Each *S* received a booklet and was instructed to complete the blank on each page of the booklet with an English word "best suited" to the words preceding it. In the case of no context, that is blank pages, *S*s were simply asked to write the first word that came to mind.

Since each booklet contained 5 of the total set of 50 contexts, responses to each of the 50 contexts were obtained from 100 *S*s. The procedure used resulted in a somewhat undesirable confounding of contexts and *S*s, but in view of the necessity for reasonably large samples it nevertheless seemed preferable to testing a sample of 10,000 *S*s that would have yielded the same amount of unconfounded data.

Results

For each of the 50 contexts, indexes of both *I* and *C* were calculated. The amount of information conveyed by the context (*I*) was computed according to Formula 1, the degree of congruity between context and target word (*C*) was given by the relative frequency of responses corresponding to the target word for each context.

Table 2 shows values of *I* and *C* for all 50 contexts. The rows represent the 10 sentences as given in Table 1, the columns indicate the length of the context, the upper number in each cell shows *I*, and the lower number shows *C*.

As can be seen from Table 2, mean *I* increases monotonically with length of context. The relation between length of context and *I* for individual contexts, however, varies greatly among contexts. In some cases there is little difference in *I* between Lengths 2 and 4 (Sentences No. 5 and 6), but in some cases the difference is quite pronounced (Sentences No. 1 and 4). Other examples of such interaction between contexts and length in determining the amount of redundancy in *S*s' responses can be easily discovered in Table 2.

Table 2 also shows that mean *C* increases monotonically with length of context. Variability in *C* among different contexts is noticeable at greater lengths of context. It is absent for Length 0 and small for Length 1 because of the fact that probabilities of correct guesses are at or near zero for these lengths.

Although both mean *I* and mean *C* increase with length of context, the two variables are relatively independent of each other. For instance, the product-moment correlation coefficient between *I* and *C* for the 20 contexts at Lengths 4 and 8 is +.439. Other lengths are ignored in this calculation since their *C* values are mostly zero or else very small.

It is obvious, of course, that the extent of the correlation between *I* and *C* is largely a matter of the selection of a sample of contexts and target words. One can easily obtain a sample in which the correlation would be very high. Given a particular sample, however, the covaria-

TABLE 2

Measures of Contextual Information (I) and Congruity between Context and Target Word (C) for 50 Contexts Used in Exp. I

Sentence No.	Length				
	0	1	2	4	8
1	.07 .00	.31 .03	.23 .02	.49 .01	.40 .33
2	.08 .00	.36 .00	.37 .01	.69 .01	.83 .05
3	.06 .00	.23 .00	.34 .07	.47 .12	.51 .36
4	.08 .00	.24 .00	.20 .05	.58 .56	.59 .56
5	.04 .00	.11 .00	.53 .19	.55 .52	.62 .59
6	.07 .00	.22 .00	.21 .00	.20 .00	.78 .79
7	.07 .00	.11 .00	.37 .00	.47 .04	.47 .21
8	.07 .00	.24 .00	.18 .00	.25 .01	.55 .01
9	.10 .00	.14 .00	.23 .00	.37 .06	.68 .43
10	.06 .00	.12 .02	.18 .02	.27 .23	.45 .13
Mean I Mean C	.07 .00	.21 .01	.28 .04	.43 .16	.59 .35

Note.—The upper number in each cell of the table shows I, the lower number shows C.

tion between VDT and I, and between VDT and C, can be determined. This was undertaken in Exp. III.

EXPERIMENT III

Method

The 20 contexts at Lengths 4 and 8 and their congruous target words, as used in Exp. I, constituted the materials of Exp. III. Fifty female students between 19 and 30 yr. of age, enrolled in various psychology courses at the University of Toronto, served as *Ss*. None had participated in Exp. I and II.

The *Ss* were randomly divided into two equal groups. VDTs for target words in presence of four-word contexts were determined for *Ss* in Group I, and those in presence of eight-word contexts were obtained from Group II. Each *S* yielded one VDT for each of the 10 target words. The order of presentation of target words was systematically varied among *Ss* to minimize any sequential effects. All other conditions of the experiment were identical with those in Exp. I.

Results

Table 3 shows mean VDTs for all 10 target words associated with four-word and eight-word contexts. Table 4 shows product-moment correlation coefficients among mean VDT, I, and C for 10 target words and for both lengths of context.

Correlation between mean VDT

TABLE 3

MEAN VISUAL DURATION THRESHOLDS FOR 10 TARGET WORDS PRESENTED IN PRESENCE OF FOUR-WORD AND EIGHT-WORD PRE-EXPOSURE CONTEXTS

Target Word	Mean Threshold (msec.)	
	Four-Word Context	Eight-Word Context
PERFORMER	80	67
COLLISION	81	80
DESERTION	69	65
SOCIALISM	44	44
RASPBERRY	52	47
AVALANCHE	88	35
BOULEVARD	68	71
GALLANTRY	79	94
INFLATION	70	63
FREQUENCY	61	73
M	69.2	63.9

Note.—Each threshold value is based on data from 25 *Ss*.

and *C* is negative and highly significant for both sets of context lengths. Other correlations are small and not significant. When the effects of *I* are partialed out, correlation coefficients between mean VDT and *C* change very little (−.918 and −.968 instead of −.928 and −.970). Partial correlations between mean VDT and *I*, with the effect of *C* eliminated, however, become −.087 and −.019 for four-word and eight-word contexts, respectively. When the data are pooled for all 20 contexts, the picture remains essentially the same: correlation between mean VDT and *C* is −.931; between mean VDT and *I*, −.349; and between *I* and *C*, +.439.

TABLE 4

PRODUCT-MOMENT CORRELATION COEFFICIENTS AMONG MEAN VDT, *I*, AND *C* MEASURES FOR THE TARGET WORDS ASSOCIATED WITH FOUR-WORD AND EIGHT-WORD CONTEXTS

	Four-Word Context		Eight-Word Context	
	I	*C*	*I*	*C*
Mean VDT	−.357	−.928	−.275	−.970
I		.352		.279

Note.—r (8) = .765 is necessary for significance at the .01 level.

Thus it seems that of the two dimensions of context, *C* accounts for a large proportion of variance in mean VDT, while *I* does not covary with mean VDT.

DISCUSSION

The results of both Exp. I and III are quite consistent. They show that length of context or the amount of information conveyed by context, per se, is irrelevant to recognition threshold of target words. The length of context, as examined in Exp. I, seems to be important only insofar as it changes the degree of congruity between context and the target word. Similarly, as shown in Exp. III, there is no covariation between VDT and contextual constraint if the effects of congruity are neutralized, but covariation is high between VDT and congruity even if contextual constraint is partialed out. Thus, it is not just the amount of information in the context that determines the ease of tachistoscopic recognition of the target word, but rather the amount of *relevant* information.

If we assume that the degree of congruity reflects the amount of relevant information available from the context, and threshold reflects the amount of stimulus information that *S* needs for correct response, then the present findings can be regarded as providing full support for the hypothesis that sources of stimulus information are interchangeable with sources of contextual information. The role of the discriminative stimulus in experiments on tachistoscopic identification of verbal material, therefore, is simply to provide information about the response to be selected that is not available from other sources. In Goldiamond and Hawkins' (1958) Vexierversuch, according to the present argument, "background" information was sufficient in many cases to provide all the information necessary for veridical response selection and *S*s were, therefore, capable of correctly "identifying" nonexistent stimulus words. Under some other conditions, such as when the forced-choice method (Blackwell, 1953) is used, nonstimulus sources of information are eliminated, and *S*'s performance is completely determined by the availability of perceptual information. In many tachistoscopic experiments in which the ascending method of limits is used to measure thresholds, however, information from several sources jointly determines the level of performance.

The findings of the present experiments are relevant to the hypothesis theory of perception (Bruner, 1951; Postman, 1951). According to this theory, perception depends upon two classes of variables—(*a*) stimulus factors, and (*b*) expectancies or hypotheses of the organism. Two basic theorems relate strength of hypothesis to perceptual information. The first says that the greater the strength of the hypothesis, the less the amount of appropriate information necessary to confirm it. The second states that the greater the strength of the hypothesis, the more the amount of inappropriate or contradictory information necessary to infirm it. Two theorems seem to be necessary to specify the relations involved since the theory is focused on hypotheses and their confirmation or infirmation. With respect to the relation between stimulus information and information from other sources, the two theorems express the same thing—the amount of stimulus information needed for veridical report is an inverse function of the amount of information available from other sources. The data from the experiments reported in this paper are in good agreement with this generalization.

REFERENCES

ABORN, M., RUBENSTEIN, H., & STERLING, T. D. Sources of contextual constraint upon words in sentences. *J. exp. Psychol.*, 1959, **57**, 171–180.

BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *U. Mich. Engng. Res. Inst. Bull.*, 1953, No. 36.

BRICKER, P., & CHAPANIS, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald Press, 1951.

BURTON, N. G., & LICKLIDER, J. C. R. Long range constraints in the statistical structure of printed English. *Amer. J. Psychol.*, 1955, **68**, 650–653.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relationship between word frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

McGILL, W. J., & QUASTLER, H. Standardized nomenclature: An attempt. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955.

MILLER, G. A., BRUNER, J. S., & POSTMAN, L. Familiarity of letter sequences and tachistoscopic identification. *J. gen. Psychol.*, 1954, **50**, 129–139.

POSTMAN, L. Toward a general theory of cognition. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951.

SMOCK, C. D., & KANFER, F. Response bias and perception. *J. exp. Psychol.*, 1961, **62**, 158–163.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

TULVING, E. Familiarity of letter sequences and tachistoscopic identification. *Amer. J. Psychol.*, 1963, **76**, 143–146.

(Received October 20, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 328–331

THE AUTOPHONIC SCALE OF VOICE LEVEL FOR CONGENITALLY DEAF SUBJECTS

HARLAN LANE

Behavior Analysis Laboratory, University of Michigan

The autophonic scale of voice level was determined by the method of magnitude production for 7 Ss with profound congenital deafness and found to be the same as that for speakers with normal hearing. Several kinds of evidence support the conclusion that the speaker does not rely on loudness in judging autophonic level and the listener does not rely on the speaker's autophonic level in judging his loudness.

The autophonic scale of voice level relates the speaker's estimate of his own vocal level to the sound pressure level produced. The reception scale of voice level relates the listener's estimate of a speaker's loudness to the sound pressure level presented. Prior research, employing several psychophysical methods for determining and validating these subjective scales, has shown that the sensory dynamics of speaking and listening are quite different. The outcome of cross-modality matching, in which *S*s equate their vocal level to sounds presented at various intensities, is predicted by the disparate exponents of the power functions that describe the autophonic and reception scales, as is the outcome of cross-modality compensation, in which *S*s adjust their vocal level, under changes in the amplification of sidetone, to maintain sidetone loudness constant. Furthermore, increased or attenuated voice sidetone, or even total masking of sidetone, has no appreciable effect on the autophonic scale (Lane, Catania, & Stevens, 1961).

These findings and others (Lane, 1962) permit the inference that the speaker does not rely on loudness in judging autophonic level and the listener does not rely on the speaker's autophonic level in judging his loudness. The present study adds further support to this conclusion by showing that the autophonic scale of voice level for congenitally deaf speakers, who have never had auditory cues available while vocalizing, is the same as that scale for normal *S*s, vocalizing with and without masking of sidetone.

METHOD

Subjects

Four female and three male, adolescent *S*s were employed;[1] each had profound binaural deafness since birth. In five cases, one or both parents was profoundly deaf, in a sixth the etiology was unknown, and in the seventh it was maternal rubella during pregnancy. The degree of hearing loss was determined by pure-tone audiometry (125–2,000 cps) and found always to exceed 70 db. and typically 100 db.

Apparatus and Procedure

The *S* sat at a distance of 18 in. from a microphone (Altec, 633 A), connected to a tape recorder (Ampex 351). All instructions were given to an interpreter who then communicated with *S* in sign language. First *S* was asked to produce a loud vowel ("very big sound"), while *E* set the recording level to zero vu. Five segments of a meterstick were then presented; the lengths of the segments were printed on their faces: 2.5, 5, 10, 20, and 40 in. The *S* was told to produce a vowel at moderate level and that the size of the sound he had just produced corresponded to the size of the middle segment ("10").

---

[1] The cooperation of the Michigan School for the Deaf, B. R. Siders, Superintendent, is gratefully acknowledged.

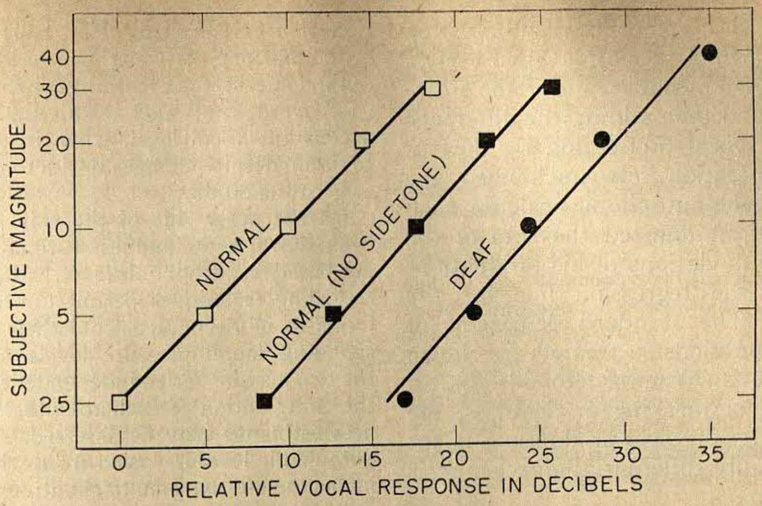


FIG. 1. The autophonic scale of voice level for seven speakers with profound congenital deafness, and for 24 *S*s with normal hearing who vocalized with sidetone and without (110 db. masking noise). (Each point is the mean decibel level of 56 vocal productions by the deaf speakers and 72 by the normal *S*s.)

Other segments were then presented in irregular order, eight times each, and the speaker was asked to respond to each with a vocal production of proportionate magnitude.

Tape recordings were processed with a graphic level recorder (Bruel and Kjaer 2305) and the relative level of each vocal response was read (in db.) from the calibrated chart.

RESULTS

Figure 1 shows the autophonic scale of voice level for the seven deaf *S*s and, for comparison,[2] that of 24 normal *S*s vocalizing with sidetone and without (110 db. masking noise). The growth of apparent vocal level as a function of the sound pressure produced is well described, for all three experimental conditions, by a straight line in log-log coordinates, in other words, a power function (Stevens, 1960). The slope (exponent) of this function, determined by a least-squares fit to the means of the db. levels produced, is much the same in the three cases: 1.2 for normal *S*s vocalizing with sidetone, 1.3 when sidetone is absent because of masking or deafness.

Table 1 permits comparison of the variability in vocal productions by the three groups of *S*s. Because the method of magnitude production was employed, the reference vocal level ("10") varied from one speaker to the next. The component of the total variance that was due to *S*'s choice of

TABLE 1

SD OF VOCAL LEVELS (in db.) IN A DETERMINATION OF THE AUTOPHONIC SCALE BY THE METHOD OF MAGNITUDE PRODUCTION

Criterion Value	Normal *S*s (Sidetone)	Normal *S*s (No Sidetone)	Deaf *S*s
2.5	3.4	3.6	4.5
5	3.3	3.4	3.2
10	2.6	3.4	2.8
20	2.8	3.0	3.4
30	4.0	3.9	—
40	—	—	4.5
S	24	24	7
N	72	72	56

Note.—The number of *S*s is denoted by *S*, the number of determinations at each criterion value by *N*.

---

[2] From Lane, Catania, and Stevens (1961).

modulus was removed by subtracting the mean of all the vocal responses by a given *S* from that mean for the entire group, and then adding the difference to each vocal production by that *S*. This operation left unchanged the slope of the autophonic scale for each *S*, but it minimized the sum of the squared deviations of his productions around the regression line for the group. The standard deviations of the productions so treated are shown in Table 1. When sidetone is absent, because of masking or deafness, the variability of vocal productions is not appreciably different than when it is present.

DISCUSSION

The finding that the form and parameters of the autophonic scale for deaf speakers are the same as that for normal *S*s confirms, in particular, that autophonic judgments need not be based on loudness and, in general, that the sensory dynamics of speaking and listening are usefully distinguished.

Our earlier research strongly supports this conclusion. First of all, the autophonic scale and the reception scale for voice level have drastically different exponents, 1.1 and 0.4, respectively (Lane, 1962). This disparity is borne out quantitatively by experiments that require *S*s to match their autophonic level to sounds of varying loudness and by complementary experiments that require *S*s to decrease autophonic level, under increases in sidetone level, to maintain sidetone loudness constant. Both kinds of equal-sensation functions, which turn out to be reciprocal power functions, have an exponent predicted by the ratio of the exponents of the autophonic and reception scales for voice level (Lane, 1962; Lane et al., 1961).

Further evidence that autophonic judgments are not based on loudness has come from experiments in which the normal relations between autophonic level and sidetone level are altered radically. One-hundred-fold amplification or attenuation of sidetone level, or total masking of sidetone, leaves the autophonic scale essentially unchanged. Moreover, the speaker readily maintains a constant vocal level, when instructed to do so, despite changes in sidetone level as great as 80 db.

On the other side of the coin, it has been shown that loudness estimates are not based on autophonic level. Magnetic recording techniques permit autophonic levels to be decoupled from their corresponding sound pressure levels so that the two may be varied independently and the contribution of each to speech-loudness estimates may be assessed. When autophonic level is varied from whispering at the one extreme to shouting at the other, the quality (spectrum) of the voice inevitably varies, but the reception scale for voice level remains invariant (Lane, 1962).

Although the conclusion seems inescapable that estimates of loudness are quite different from estimates of autophonic level, Warren (1962) has recently contended, on the contrary, that they are the same, and that both kinds of estimates are based on spatial localization. Observers learn to judge the loudness of externally produced sounds, he suggests, by noting how their loudness changes with distance. Self-generated sounds do not, of course, vary in distance, but the speaker has learned, so the theory goes, how much he must increase his loudness to reach listeners at varying distances.

The "physical-correlate theory" (Warren, 1958) has been criticized previously (Stevens, 1959). This more recent extension of the theory, relating distance, loudness, and autophonic level, seems faulty in all three correspondences. First, the prediction that loudness grows linearly with sound pressure, which follows from the inverse-square law relating distance to sound intensity, is counter to the findings of innumerable studies.[3] Second, the contention that estimates of autophonic level and of loudness are identical is counter to the several findings cited above. Warren disputes the conclusiveness of one of these findings, namely that the autophonic scale is invariant under total masking of sidetone. He suggests that the speaker's familiarity with the relation between autophonic and sidetone level provides a basis for loudness judgments even in the absence of auditory cues. This explanation of the invariance of the autophonic scale cannot apply, however, to the congenitally deaf *S*s who, lacking this familiarity, nevertheless provided the same subjective scale as did normal speakers.

---

[3] The most recent study is by Stevens and Guirao (1962).

Finally, the hypothesis that a speaker's estimate of his autophonic level is in fact an estimate of his distance from a listener, on those prior occasions when he spoke at that loudness, notably lacks support in the present confirmation of the autophonic scale with congenitally deaf *S*s.

The sensory basis for autophonic judgments is more likely to be found among the cues to vocal effort. Ladefoged and McKinney (1963) report that the rate of work done upon the air in producing a voiced sound is proportional to the 1.2 power of the sound pressure. Since the autophonic scale has the same form and exponent as this vocal work function, the authors suggest that estimates of autophonic level are based on physiological effort—which is what our *S*s have been telling us for some years now.

REFERENCES

Ladefoged, P., & McKinney, N. P. Loudness, sound pressure, and subglottal pressure in speech. *J. Acoust. Soc. Amer.*, 1963, **35**, 454–460.

Lane, H. L. Psychophysical parameters of vowel perception. *Psychol. Monogr.*, 1962, **76**(44, Whole No. 563).

Lane, H. L., Catania, A. C., & Stevens, S. S. Voice level: Autophonic scale, perceived loudness, and effects of sidetone. *J. Acoust. Soc. Amer.*, 1961, **33**, 160–167.

Stevens, S. S. On the validity of the loudness scale. *J. Acoust. Soc. Amer.*, 1959, **31**, 995–1003.

Stevens, S. S. The psychophysics of sensory function. *Amer. Scientist*, 1960, **48**, 226–253.

Stevens, S. S., & Guirao, M. Loudness, reciprocality, and partition scales. *J. Acoust. Soc. Amer.*, 1962, **34**, 1466–1471.

Warren, R. M. A basis for loudness judgments. *Amer. J. Psychol.*, 1958, **71**, 700–709.

Warren, R. M. Are "autophonic" judgments based on loudness? *Amer. J. Psychol.*, 1962, **75**, 452–456.

(Early publication received May 15, 1963)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 332–337

INTERACTION OF NOISE WITH KNOWLEDGE OF RESULTS AND SLEEP DEPRIVATION[1]

ROBERT T. WILKINSON

Applied Psychology Research Unit, Cambridge, England

100 db. "white" noise impairs performance of 30-min. choice serial reaction; does this effect vary under (Experiment I) 32-hr. sleep deprivation (SD) and (Experiment II) increased knowledge of results (KR)? 12 enlisted men in each experiment carried out the 4 condition combinations in balanced order of presentation. The effect of noise was increased by KR, reduced by SD, and greater among Ss with previous practice on the test. Conclusions are: (a) noise impairs performance as incentive is high and as the task loses novelty through practice; (b) noise and SD produce different types of "fatigue" which may oppose each other's action; with SD arousal may be too low (especially with no KR); with noise it may be too high (especially with KR).

These two experiments are reported from part of a protracted series on the interaction of stresses and related variables, the aim of the program being to learn more about these influences by observing how the impact of one changes in the presence of another. As far as possible the experimental design, the *S*s, and the general atmosphere of the experiments are kept constant from one study to the next. In particular the same test is always used, namely Leonard's 5-Choice Test of Serial Reaction (Leonard, 1959). Previous experiments have dealt with the interaction of sleep deprivation (SD) with rest pauses during the test (Wilkinson, 1959), with a warm environment (Pepler, 1959), and with knowledge of results (KR) (Wilkinson, 1961). The latter study involving SD and KR is particularly relevant because we are now concerned with the influence of these two variables upon the known (Broadbent, 1953) effect of continuous "white" noise (100 db.) on the same test.

METHOD

Subjects.—The *S*s were enlisted men between the ages of 18 and 30. Twelve took part in Exp. I and a further 12 in Exp. II. All possessed normal hearing.

Apparatus.—The test has already been fully described in the reports of previous experiments in the series (Wilkinson, 1959, 1961). Briefly, there are five light bulbs and five associated metal contacts. The *S* holds a stylus and taps the contact whose bulb is lighted. This extinguishes the light and brings another on, which is again responded to; and so the cycle continues with *S* working at his own pace for 30 min. The order in which the bulbs light appears random. There are three main scores. The first is the number of correct responses (corrects). If an incorrect contact is touched the stimulus still changes but an error is scored, and finally a gap is counted whenever *S* allows a period of 1½ sec. to elapse between responses.

Procedure.—The *S* worked alone in a partially soundproof cubicle.

The noise was continuous open field "white" noise at 100 db. intensity measured at the ear of the *S* by a Dawe Type 1400 E sound level meter. The noise was recorded on tape and played back through a standard Ferrograph recorder and a suitably mounted 12-in. Whiteley electrical speaker. In the quiet condition this noise was removed leaving only a slight residual speaker noise and faint room noises which occasionally filtered through the walls of the cubicle.

---

[1] The Royal Navy supported this work by contributing *S*s, administrative help, and financial assistance through the Royal Naval Personnel Research Committee of the Medical Research Council of the United Kingdom. The author wishes to thank D. E. Broadbent for his helpful advice; and R. C. Collecot and P. Grant-Ross, Chief Petty Officers, and R. Edwards, Sick Berth Attendant, for administering the tests.

The form of sleep deprivation (SD) has again been described before. The *S*s remained awake the previous night until tested during the afternoon of the next day. This amounted to about 32 hr. without sleep. During this time they played games, read books, and carried out routine duties always under supervision. With normal sleep the men carried out exactly the same routine except that they were allowed to sleep normally the previous night.

Knowledge of results (KR) was given in three ways. Immediate information was provided by two peripheral lights. One flashed whenever an error was made and the other signaled gaps by lighting as soon as 1½ sec. had elapsed since the previous response. Every 5 min. a note was placed before *S* giving scores of corrects, errors, and gaps for the previous 5 min. and, after the first 5 min. relating these scores to those of the 5 min. before by the words "Worse," "Better," or "No change." Finally the totals of corrects, errors, and gaps were published for all *S*s to see. In the normal form of the test with no knowledge of results (NKR) none of these forms of feedback was provided and *S*s were told of this at the outset.

Design.—In Exp. I *S*s were tested on the Tuesday and Thursday of 2 successive wk. and always at the same time of day. They met each of the four possible combinations of noise and quiet with normal sleep and SD in a 4 × 4 Latin square design which balanced order effects. The procedure was just the same for *S*s of Exp. II except that they met each possible combination of noise and quiet with KR and NKR.

Practice.—On the Monday before the first main test *S*s were given ½-hr. practice which acquainted them with the forms of the test they were to receive. In Exp. II this was the only practice given, but in Exp. I *S*s had a total of some 3 hr. previous experience of the test as part of other experiments involving SD, intermittent aircraft noise (which had no statistically significant effect), and the variable of sitting or standing at the test. Administratively this difference in practice in the two experiments was unavoidable. Its importance will emerge later.

Statistical treatment.—The significance of means is tested by Wilcoxon's nonparametric procedure as many of the distributions were not normal. The Mann-Whitney *U* test is used to assess differences between means, and Kendall's tau to examine correlations. All these procedures are described in Siegel (1956). Two-tailed levels of significance are given except where stated otherwise.

RESULTS

Experiment I: Noise and Sleep Deprivation

Totals of gaps, corrects, and errors have been calculated for each 10 min. of the test and for each of the four possible combinations of sleep deprivation (SD) and normal sleep with noise and quiet. These scores are shown averaged over all *S*s in Fig. 1. The most pronounced differences between conditions occurred towards the end of the test and therefore the analysis will be confined to scores over the last half of the 30-min. period.

The effect of SD in both noise and quiet was to reduce corrects ($p < .01$) and increase gaps ($p < .01$) (Fig.1). Also, as before, SD had no significant effect upon the absolute number of errors in either noise or quiet (Fig. 1). However this index may become unrealistic in experiments involving loss of sleep if, in the limit, *S* falls asleep for some time and, making no responses, achieves the rather flattering score of zero in errors. For this reason another index of accuracy has been derived, proportional errors, which represents the proportion of responses made which are incorrect as a percentage $\left(\frac{E \times 100}{E + C}\right)$. The adverse effect of SD on this index of accuracy was even more marked than upon gaps and corrects (Fig. 1). It appeared in 11 out of 12 *S*s under quiet ($p < .001$) and in 8 out of 11 *S*s in noise ($p < .05$). It will be noted that the effect of SD was less under noise. This interaction will be considered in more detail in the context of noise to which we now turn.

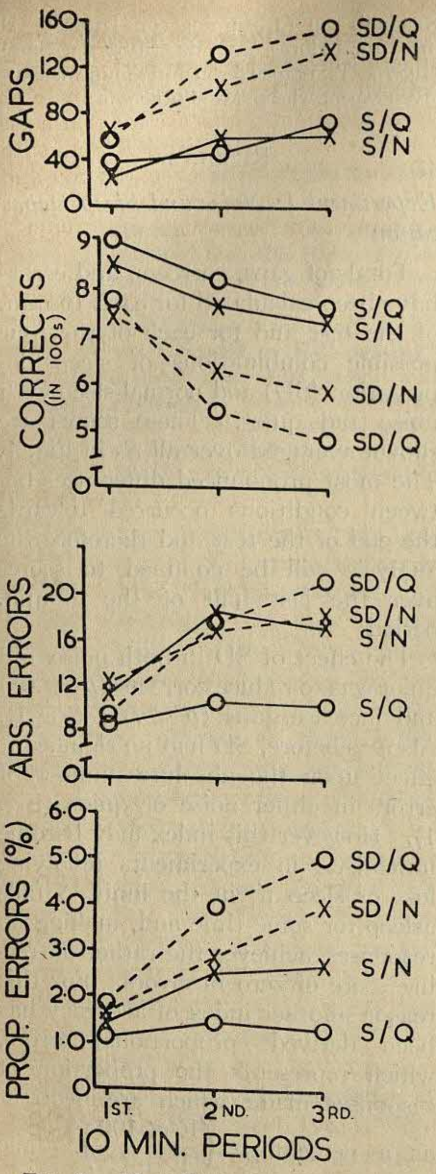


FIG. 1. The effect of combinations of sleep (S) and sleep deprivation (SD) with noise (N) and quiet (Q) upon: gaps, corrects, absolute errors, and proportional errors (errors as a percentage of all responses).

In terms of both gaps and corrects there was no significant effect of noise upon performance either with or without sleep. The same holds for errors in the sleep and SD conditions combined, but a different picture emerges when these conditions are considered separately. Noise after normal sleep increased the number of both absolute and proportional errors in 11 out of 12 *S*s ($p < .01$ in each case) but had no significant effect under SD. Thus SD reduced the adverse influence of noise. This interaction was significant in absolute errors ($p < .05$) but even more marked in proportional errors where 11 out of 12 *S*s showed a reduced effect of noise when sleep had been lost ($p < .01$).

Experiment II: Noise and KR

Figure 2 shows scores of gaps, corrects, and errors arranged as before but now in the four combinations of noise and quiet with KR and NKR. Performance as a whole was better in this experiment than in the first. In contrast to Exp. I the most marked results appeared not in errors but in gaps, where different effects of noise occurred in NKR and KR. In NKR gaps were not significantly affected by noise; with KR they were increased ($p < .02$). This interaction was highly significant ($p < .01$), 10 out of 12 *S*s showing a greater adverse effect of noise when KR was added. The same interaction can be restated in terms of KR: this feedback reduced gaps more in quiet ($p < .01$) than in noise ($p < .05$, one-tailed). Similar results appeared in errors, although with less clarity and significance. Noise impaired accuracy with KR ($p < .05$, two-tailed in proportional errors, one-tailed in absolute errors) but had no effect under NKR. This interaction ($p < .05$, one-tailed) can be restated in terms of KR: this feedback improved accuracy more in quiet ($p < .05$) than in noise. As before little of significance appeared in corrects. To sum up, in gaps and errors, but not in corrects, the adverse effect of noise was increased with feedback of KR, and, complementarily, as noise was increased KR became less effective in improving performance. Finally in this section a comparison is made across experiments: The adverse effect of noise on accuracy was greater in the sleep condition of Exp. I than in the NKR condition of Exp. II ($p < .01$ in both absolute and proportional errors) (Fig. 1 and 2). The two conditions appear equivalent except for the greater preliminary practice involved in Exp. I.

DISCUSSION

Earlier findings confirmed.—In the normal form of the 5-Choice Test the following results have been confirmed: (*a*) KR improves all aspects of performance (Wilkinson, 1961). (*b*) SD greatly increases gaps and also reduces corrects (Pepler, 1959; Wilkinson, 1959, 1961). As in previous studies SD did not significantly increase the absolute number of errors, but it did reduce the number of responses made so that the *proportion* of errors rose with SD. Accuracy, when analyzed on this basis, proves at least as sensitive to SD as gaps and corrects. (*c*) Broadbent's (1953) finding that 100 db. "white" noise increased errors towards the end of the test was confirmed under the normal (sleep) condition of Exp. I but not in the corresponding normal condition (NKR) of Exp. II. The only apparent difference between the two settings was that *S*s of Exp. I had considerably more preliminary practice on the task. We can only conclude that this greater familiarity of the task was what made it more vulnerable to noise in the first experiment. The implication of this conclusion is that the effect of noise on repetitive work may increase as the task loses its novelty and the power of absorbing attention. In a similar setting (Wilkinson, 1961) a similar conclusion has been reached concerning the effects of SD.

Interaction of noise, SD, and KR.—The adverse effect of noise on performance is reduced by SD and increased by KR; the implications of this are (*a*) that noise has more in common with KR than with SD, and (*b*) that the effects of SD and KR oppose each other, which confirms the previous finding (Wilkinson, 1961).

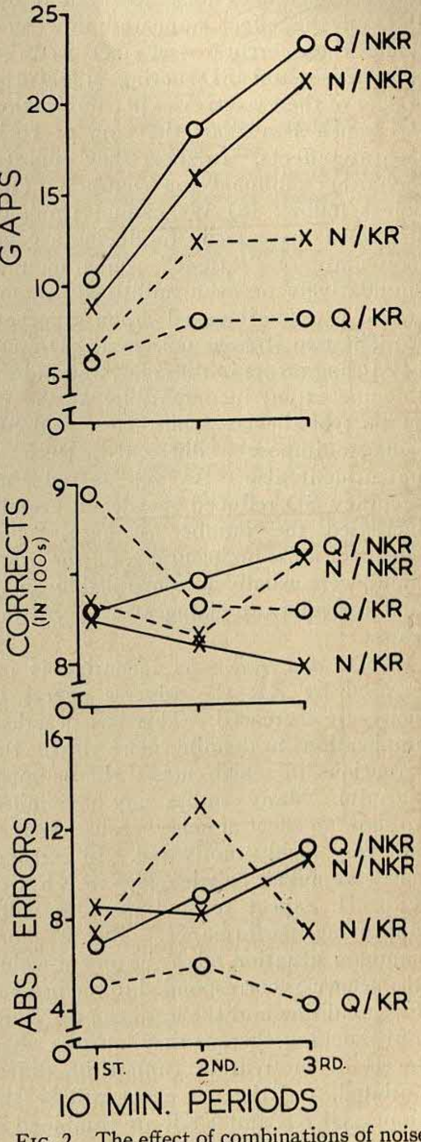


FIG. 2. The effect of combinations of noise (N) and quiet (Q) with knowledge of results (KR) and no knowledge of results (NKR) upon: gaps, corrects, and absolute errors.

Noise and SD.—"White" noise at 90 db. through headphones has been found to reduce gaps in choice serial reaction time under sleep deprivation but not with normal sleep (Corcoran, 1962). This suggested a negative interaction between the two stresses and this has been demonstrated in the present experiment. Here however open field noise was used at 100 db. and the effect appeared in errors. While these errors were increased by both noise and SD acting singly the impact of the two stresses in combination was much less than the sum of their separate effects. In short they opposed each other's influence on accuracy. Two points follow: (*a*) We cannot think of stresses as necessarily producing a common state of "fatigue" which mounts cumulatively or even additively as one stress is superimposed upon another. (*b*) The two stresses noise and SD must be causing errors in different, and indeed to some extent incompatible, ways. In support of the last point noise and SD seem to impose a different pattern of impairment also. As well as reducing accuracy SD reduced speed and greatly increased the number of gaps. The effects of noise on accuracy, on the other hand, were usually accompanied by small and insignificant changes in speed and gaps.

Noise and KR.—As incentive is increased by KR the adverse effects of noise are increased. This has practical implications in defining more clearly the situations in which noise will be most harmful. Many studies may have failed to show an effect of noise because *S*s were not very highly motivated. In extreme cases we might, as with gaps in NKR of Exp. II, expect to find noise actually improving performance. But a more common situation might be one in which the keener *S*s corresponded to the present KR condition and the less keen to NKR. Put them together, as they must be when levels of motivation cannot be differentiated, and the net effect will be the same as if KR and NKR are combined in Exp. II (gaps), that is no net effect of noise but a high individual variability. This is a familiar result and one which has often been taken to indicate no effect of noise.

Role of KR.—It is strange to find KR, normally a benevolent influence, working in league with noise to debase performance, but following Elwell and Grindley (1938) we must distinguish two functions of KR: First, it will give *S* information on the effectiveness of his responses, and second, it may increase the penalties and rewards attached to his performance. It is difficult to see how the former could impair performance, but the latter, obeying the Yerkes-Dodson law, might do so under certain conditions. In such a case the overall advantage of KR would be reduced due to the disadvantages of excessive rewards or punishment offsetting the benefit of the informational feedback. Now Yerkes and Dodson (1908) noticed that the tendency for punishment to become too great for optimal learning increased with the difficulty of the task; and we have seen that KR became less effective under noise (Exp. II) and more so under SD (Wilkinson, 1961). Is there some connection, and if so what is the link between Yerkes' "difficulty" and the present stresses of noise and SD? A possible answer is that what has become known as arousal may be the intervening variable which not only underlies this parallel but may also link a number of the trends we have been observing. Evidence from physiological "arousal measures" suggests that SD lowers arousal (Armington & Mitnick, 1959; Ax & Luby, 1961; Bjerner, 1949; Wilkinson, 1962) and that noise (Helper, 1957; Ryan, Cottrell, & Bitterman, 1950) and KR raise it (Stennett, 1957; Wilkinson, 1962). Thus combinations of SD, NKR, and quiet should be minimally arousing and those of noise with sleep and especially with KR maximally so. The fact that we have found performance to be impaired at both of these extremes harmonizes with the inverted ᴜ model which is supposed to relate performance and arousal (Duffy, 1934; Freeman, 1948; Hebb, 1955; Lindsley, 1951; Malmo, 1959).

REFERENCES

ARMINGTON, J. C., & MITNICK, L. L. Electroencephalogram and sleep deprivation. *J. appl. Physiol.*, 1959, **14**, 247–250.

AX, A., & LUBY, E. D. Autonomic responses to sleep deprivation. *Arch. gen. Psychiat.*, 1961, **4**, 55–59.

BJERNER, B. Alpha depression and lowered pulse rate during delayed actions in a serial reaction test. *Acta physiol. Scand., Stockholm*, 1949, **19** (Suppl. No. 65).

BROADBENT, D. E. Noise, paced performance and vigilance tasks. *Brit. J. Psychol.*, 1953, **44**, 295–303.

CORCORAN, D. W. J. Noise and loss of sleep. *Quart. J. exp. Psychol.*, 1962, **14**, 178–182.

DUFFY, E. Emotion: An example of the need for reorientation in psychology. *Psychol. Rev.*, 1934, **41**, 184–198.

ELWELL, J. L., & GRINDLEY, G. C. The effect of knowledge of results on learning and performance. *Brit. J. Psychol.*, 1938, **29**, 39–53.

FREEMAN, G. L. *The energetics of human behavior.* Ithica: Cornell Univer. Press, 1948.

HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–254.

HELPER, M. M. The effects of noise on work output and physiological activation. *USA Med. Res. Lab. Rep.*, 1957, No. 270.

LEONARD, J. A. 5 choice serial reaction apparatus. *Med. Res. Council Appl. Psychol. Res. Unit Rep., Great Britain*, 1959, No. 326/59.

LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

PEPLER, R. D. Warmth and lack of sleep: Accuracy or activity reduced. *J. comp. physiol. Psychol.*, 1959, **52**, 446–450.

RYAN, T. A., COTTRELL, C. L., & BITTERMAN, M. E. Muscular tension as an index of effort: The effect of glare and other disturbances in visual work. *Amer. J. Psychol.*, 1950, **63**, 317–341.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

STENNETT, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–61.

WILKINSON, R. T. Rest pauses in a task affected by lack of sleep. *Ergonomics*, 1959, **2**, 373–380.

WILKINSON, R. T. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. *J. exp. Psychol.*, 1961, **62**, 263–271.

WILKINSON, R. T. Muscle tension during mental work under sleep deprivation. *J. exp. Psychol.*, 1962, **64**, 565–571.

YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit-formation. *J. comp. neurol. Psychol.*, 1908, **18**, 459–482.

(Received September 21, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 338–346

RESPONSE MEDIATION OF THE CONDITIONED EYELID RESPONSE[1]

G. ROBERT GRICE AND JOHN J. HUNTER

University of Illinois

Employing the Grice and Davis (1958) paradigm for response-mediated generalization, 3 experiments were conducted in which an air puff UCS was presented by S's mediating response. Evidence for response-mediated generalization was obtained when the UCS was an immediate consequence of the response, but the effect was greater if the response was manual than if it was verbal. Mediated, eyelid responses were roughly synchronized with the mediating reaction. When the UCS followed the mediating response by .5 sec., to insure a favorable CS-UCS interval with respect to response-produced cues, no mediated generalization was obtained. While these experiments provide evidence for a form of response-mediated generalization, they do not support a response-produced-cue theory of the phenomenon.

Evidence for the response mediation of a generalized, conditioned eyelid response has been presented by Grice and Davis (1958, 1960). A CS, always paired with an air puff UCS, and two negative stimuli, never reinforced, were presented in a differential conditioning arrangement. The *S*s also made two instructed, manual reactions to the three stimuli. One of these was made in common to the CS and to one of the negative stimuli. The other was made to the second negative stimulus. Evidence for the mediational effect consisted of a greater number of eyelid responses to the negative stimulus having the manual response in common with the CS than to the negative stimulus leading to the manual response not so associated. In the experiments producing such evidence, the mediating reactions were overt, discrimination reaction time responses to the three stimuli; and occurred within the CS-UCS interval, and in a very short period beyond. Thus, the mediating response and the eyelid response always occurred in close temporal proximity. In experiments in which the mediating reactions were not overt, or in which they occurred consistently before or after the CS-UCS interval, results have been negative (Grice, Simmons, & Hunter, 1963).

In view of the obvious sensitivity of this phenomenon to temporal relationships, as well as the importance of such variables in conditioning generally, it is apparent that the time relations require more extensive and precise exploration. One problem involved in such experimentation with this situation is the inherent variability of the reaction time of the mediating response. The variable reaction time itself constitutes the interval from the CS to the mediating response. With a fixed CS-UCS interval, the temporal relation between the response and the UCS-UCR sequence is also variable. In the Grice and Davis experiments, the response occurred usually at a variable interval before the UCS, but, fairly frequently, shortly after it. In the present experiments, this source of variability was removed by arranging for the UCS, an air puff, to be presented by the occurrence of the mediating response. Miller and Cole (1936) reported an eyelid conditioning experiment in which the UCS, shock, was administered by *S*'s pressing a telegraph key in response to a CS. This procedure resulted in an extremely rapid and stable increase in anticipatory blinks. However, in view of the likelihood of facilitative effects in this situation, Grice and Davis (1960) raised a question concerning the extent to which conditioning or associative processes are actually involved in this increase. The differential conditioning procedure of the present series would appear to be preferable to the Miller and Cole simple conditioning situation for the use of this technique. Here, a response is made to each of the two negative stimuli, but only the one in common with the CS is ever paired with the UCS. Thus, any purely facilitative effect should be equal for the two stimuli. Any difference between the two stimuli in frequency of eyelid responses may be attributed to the pairing of one response with the UCS; hence, such a difference must be viewed as associatively determined. Any differential effects specific to the particular mediating responses or stimuli employed are controlled by complete counterbalancing of these variables. Experiment I, reported here, is of this design. The CS was a light, and the negative stimuli were tones. Mediating reactions were the two lever pushing responses used in the previous experiments—performed as RT responses to the three stimuli. In the case of the CS only, the lever response delivered the air puff.

---

[1] The present research was supported by Grant G-14223 from the National Science Foundation.

Experiment II was identical to Exp. I except that the mediating responses were words spoken by *S* as RT responses to the stimuli. Grice, Simmons, and Hunter (1963) failed to obtain evidence for mediational effects of verbal responses, but it appeared likely that the negative results were attributable to an unfavorable temporal relation between these responses and the UCS. The response-contingent UCS should insure a more consistent and favorable relationship.

EXPERIMENTS I AND II

Method

Subjects

The *S*s were undergraduate women who served as part of a requirement for a course in introductory psychology. There were 24 *S*s in each experiment.

Apparatus

The general procedures and apparatus have been previously described (Grice & Davis, 1960). The apparatus included electromechanical recording equipment and programing equipment for the automatic presentation of 100 trials. The CS was a light (Grice & Davis, 1960, Exp. 2). The negative stimuli were tones of 240 and 1,900 cps, presented at approximately 70 db. loudness level. In Exp. I the mediating responses were the right- and left-hand lever switch, snapping responses of the previous experiments. In Exp. II, the latencies of the spoken words used as mediating responses were recorded by means of a voice key operated by a throat microphone. The UCS was a puff of air of 1 psi. On trials when the CS was presented, the UCS was delivered by operation of the switch in Exp. I, or by means of the voice key in Exp. II. Due to lag in the air line, there was approximately a 60-msec. interval before arrival of the air puff at the eye. The light and tones were terminated by the switch or by the vocal response.

Responses scored were those resulting in a pen deflection of 1 mm. or more with the gain setting employed. On a very small number of trials there were two scorable responses. In this case, the response recorded was that occurring nearest to the mediating response. Typically, the second response was either of very short latency, i.e. about 150 msec., or of very long latency, about 800 msec., following the mediating response.

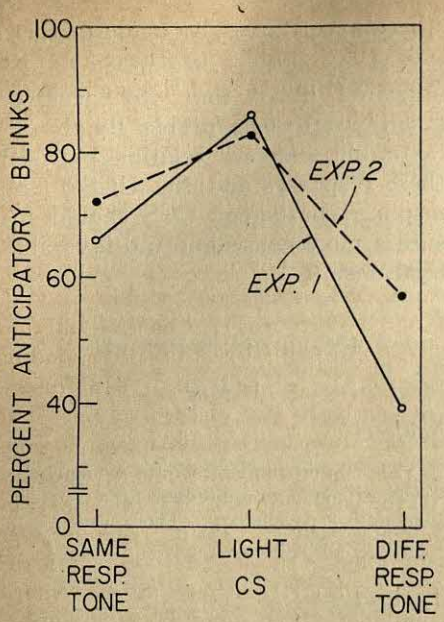


FIG. 1. Percent of anticipatory eyelid responses for Exp. I and Exp. II.

Procedure

Experiment I.—The *S* was instructed to make one of the lever snapping responses as quickly as possible to the light and to one of the tones, and the second lever response to the other tone. The UCS was delivered by the switch response to the light only. The three stimuli were presented in an irregular order for a sequence of 100 trials. Trials 41–100 consisted of 20 presentations of each stimulus. Altogether there were 40 conditioning trials with the CS, and 30 presentations of each negative stimulus. There was complete counterbalancing of the two tones and of the right and left manual responses, resulting in four counterbalancing conditions as described by Grice and Davis (1958).

TABLE 1

NUMBER OF BLINKS AND REACTION TIMES TO NEGATIVE STIMULI

Stimulus	Number of Blinks				RT in Msec.			
	Exp. I		Exp. II		Exp. I		Exp. II	
	M	*SD*	*M*	*SD*	*M*[a]	*SD*	*M*[a]	*SD*
Same Response	15.2	4.87	15.8	4.91	655	129	814	177
Diff. Response	11.0	6.80	13.7	5.78	636	114	848	158

[a] Mean of *S* means.

Experiment II.—The experiment is the same as Exp. I, except that verbal responses were used in place of the switch responses. For example, in one of the conditions, *S* was instructed to say the word "One" as quickly as possible to the light and to the high tone, and the word "Two" to the low tone. The response "One" was common to the light and a tone for half of the *S*s, and the word "Two" for the other half. The high and low tones were similarly counterbalanced, making a total of four conditions as in Exp. I. On CS trials, operation of the voice key terminated the light and presented the puff. On negative trials it terminated the tone. Latency of the verbal response was recorded on the oscillograph.

TABLE 2

ANALYSES OF NUMBER OF BLINKS TO NEGATIVE STIMULI

Source	*df*	*F*		
		Exp. I	Exp. II	Exp. III
Between *S*s	23			
Tone arrangement (T)	1	1.53	.33	.004
Response arrangement (R)	1	8.08*	.16	2.29
T × R	1	.08	3.14	.44
Error (b)	20	(49.1)[a]	(45.8)[a]	(90.88)[a]
Within *S*s	24			
Generalization (G)	1	23.42***	5.09*	.06
G × T	1	.00	1.33	.24
G × R	1	2.70	2.60	.06
G × R × T	1	2.10	.03	.06
Error (w)	20	(8.90)[a]	(9.83)[a]	(5.48)[a]
Total	47			

[a] Error *MS*s.
* $p < .05$.
*** $p < .001$.

Results

Data presented are numbers of eyeblinks on Trials 41–100, the block including 20 presentations of each stimulus. The mean percentages of anticipatory blinks are presented in Fig. 1. Anticipatory blinks are defined as those occurring from 150 msec. following onset of the CS through the latency period of the UCR, 80 msec. following completion of the mediating response. This includes the lag time of the air system. As predicted, there were more responses to the tone having the response in common with the CS in both experiments. However, the asymmetry is substantially greater with the manual responses of Exp. I than with the verbal responses of Exp. II. Since no UCS accompanied the negative stimuli, responses to them were recorded for the interval 150 msec., following the CS onset until 800 msec. following the mediating response. Means of these data are presented in Table 1. Again, the predicted asymmetry is observed in both experiments but is greater for Exp. I. Analyses of variance of these data are presented in Table 2. Counterbalanced variables appear as between-*S*s effects and the test of asymmetry is the Generalization effect within-*S*s. The effect is significant in both experiments.

Possible effect of differential latency of the mediating response upon generalization of the eyeblink should be evaluated. If there were a systematic tendency to respond faster to the Same Response stimulus, and if fast responses were associated with a greater probability of a blink, the asymmetry in frequency of eyeblinks could be attributed to a nonassociative process. In the present frame of reference, the result would be regarded as artifactual. However the data do not support such an interpretation. Means of *S* mean RTs are presented in Table 1. Differences between RTs to the two stimuli are small and in opposite directions in the two experiments. Neither difference is significant (Exp. I: $t = 1.45$, $p > .10$; Exp. II: $t = -1.52$, $p > .10$). Furthermore, there is no evidence for association between latency of the mediating response and probability of an eyelid response. For each experiment, and for each stimulus separately, scatter plots were made of the relation between *S* mean latency and the number of blinks. All of these plots obviously indicated correlations of zero order. In Exp. I, there were 19 *S*s who made neither 0 nor 100% blinks to the Same Response stimulus and 18 such *S*s for the Different Response stimulus. For these *S*s, the mean latency of the switch response was computed for blink and nonblink trials separately. For the Same Response stimulus the means were 669 msec. on blink trials and 682 msec. on nonblink trials ($t = 0.39$, $p > .7$). For the Different Response stimulus, the corresponding means were 642 and 648 msec., respectively ($t = 0.28$, $p > .7$). The implication of these last comparisons is that there is no within-*S* dependency of the probability of a blink upon the speed of the mediating response. In summary, these analyses indicate, first, that there is no systematic tendency to respond faster to either of the two negative stimuli. They also indicate, further, that there is neither a between-*S* nor a within-*S* relation between the speed of the mediating response and the probability of an eyeblink. Apparently, the obtaining of these latter correlations by Grice and Davis (1960) was due entirely to their use of a brief, fixed CS-UCS interval. Only fast responses had an opportunity to produce blinks within the scoring interval.

Of considerable importance in the interpretation of the mediational effects obtained here is the temporal relation between the mediating response and the eyeblink on trials when both occur. Frequency distributions of this relationship for Exp. I are presented in Fig. 2. The zero point on this graph represents the moment of contact of the switch, i.e., the completion of the switch closing response. Negative values plotted to the left indicate intervals before this time, and positive values to the right indicate intervals following the switch response. It will be noted that, in the case of all three stimuli, the eyelid

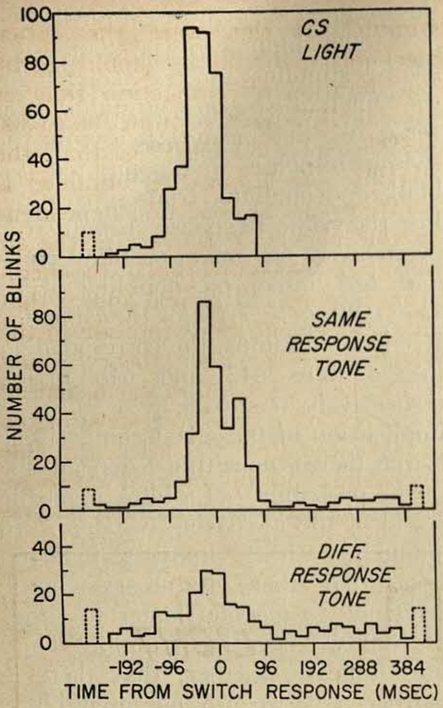


FIG. 2. Distributions of latencies of eyelid responses in Exp. I plotted with respect to completion of the switch response. (End categories enclosed by dotted lines indicate scattered responses in the tails of the distributions.)

response has a modal frequency within a few milliseconds of the closing of the switch. In the case of the CS, there is slightly more of a tendency for the blink to lead the manual response. The general picture appears to be one of approximate synchronization of the two responses, and agrees roughly with the findings of Miller and Cole (1936). The degree of synchronization and the general precision of these data are indicated by the fact that the interquartile range of the distribution for the CS is only 55 msec. The nature of the mediational effect is suggested by a comparison of the distributions for the two negative stimuli. Since the difference appears to be primarily in the size of the mode, it appears that the asymmetry is produced by a greater tendency to synchronize blinks with the Same Response than with the Different Response.

It is possible to plot the same data with respect to the time of stimulus onset rather than the time of the manual response. This has been done in the case of the negative stimuli. While not presented, these distributions are much more platykurtic, tending to rectangularity. The interquartile ranges of the data of Fig. 2 are 80 msec. for the Same Response and 217 msec. for the Different Response. The corresponding values for the same data plotted from the stimulus onset are 274 and 380 msec., respectively. There can be little doubt that the blinks are predominately under the control of the mediating response rather than of the auditory stimulus.

Similar latency data for Exp. II are presented in Fig. 3. The result is similar, but the modes are broader and there is little evidence of a clearly defined mode in the case of the Different Response. There is also a tendency for the blink to lead the operation of the voice key by a substantially larger time than the corresponding interval in Exp. I. Speculation concerning the reason for this is probably not very profitable, since it is not known at what stage in the formation of the response the word becomes sufficiently voiced to operate the key. Again, however, it appears that the asymmetry obtained is produced by blinks closely related to the Same Response.

EXPERIMENT III

While Exp. I and II provide rather convincing evidence for response-mediated generalization, it appears quite unlikely that its mechanism could have been the one usually assumed to produce it. Theoretical accounts of mediated generalization, or mediated stimulus equivalence, have been in terms of response-produced cues which are assumed to elicit the generalized response. If some response is made in common to two stimuli, it is the cues produced by this response which are supposed to result in increased equivalence of the stimuli. While measurements are not available to describe fully the temporal relations of the responses in the above experiments, it is fairly clear that the eyeblinks occurred too close to the mediating responses to be dependent on movement-produced stimuli. If this were the actual mechanism, the blinks should lag by the duration of a reaction time or typical CR latency of from 200 to 500 msec. One possibility is that the response synchronization obtained is a rather special condition dependent on the immediate delivery of the air puff with the completion of the mediating response. It would appear that an arrangement more apposite to the response-produced-cue theory would be for the UCS to follow the mediating response by a time corresponding to a favorable CS-UCS interval. One approximation of such an arrangement was attempted without success

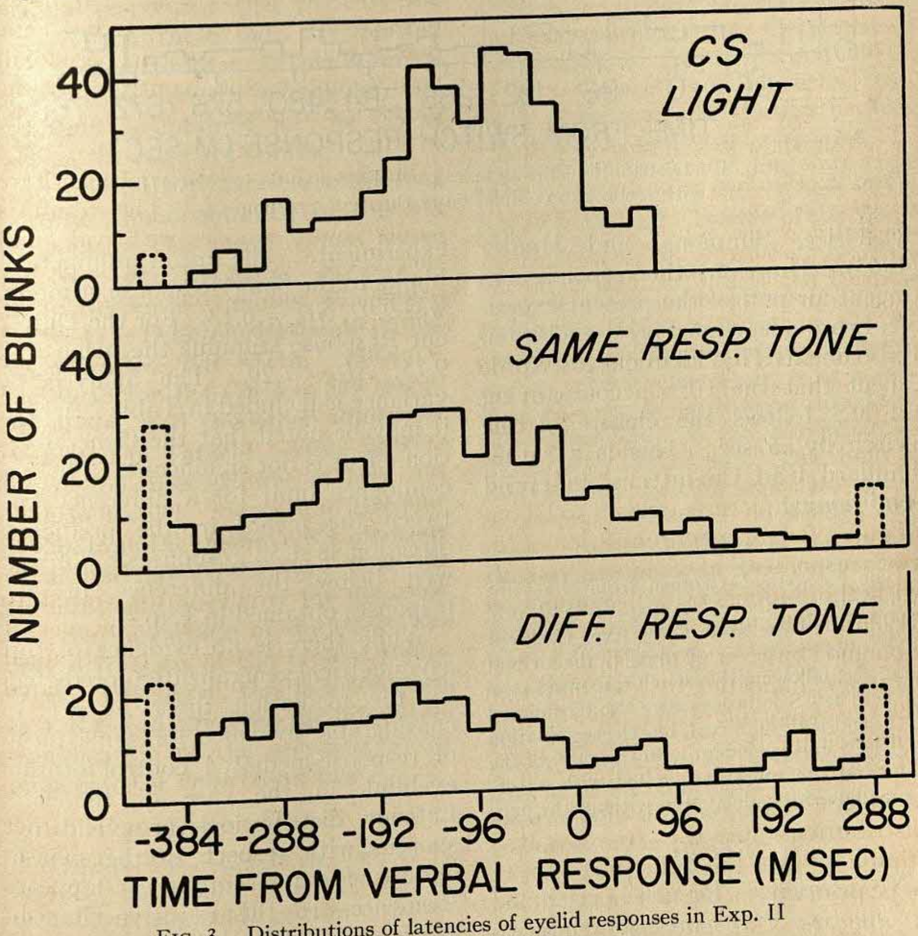


FIG. 3. Distributions of latencies of eyelid responses in Exp. II with respect to the verbal response.

tained. In Exp. I, particularly, the asymmetrical generalization appears to be produced entirely by a modal frequency of blinking in the immediate region of the mediating response. This close temporal proximity seems to rule out mediation on the basis of response-produced cues from the manual response. Two possible alternative mechanisms are suggested. One is that both are mediated by some common, short latency, central mediating process. But if this is true, the mechanism by which this process generalizes from the light to one tone more than the other is not clear. The second suggestion concerns the possible operation of a principle of response integration. There may exist a form of associative, R-R learning by means of which two or more responses become organized into a coordinated act. Generalization could then occur when a stimulus for one of the responses evokes the entire integrated response system. This would be a mechanism for response-mediated generalization quite different from those usually discussed. The presence of a reduced, but clearly defined, modal frequency of blinks to the Different Response tone suggests that such integrated response systems may be subject to bilateral transfer. While Grice and Davis (1960) attributed the production of blinks by the Different Response to purely facilitative effects, the absence of such findings with the delayed puff of Exp. III suggests that their interpretation may be incorrect. This interpretation in terms of response integration must be regarded as very tentative pending further investigation of the parameters involved in such situations. Its significance as a general principle would also depend upon its demonstration with response systems other than the eyeblink.

The interpretation of the results of Exp. III is somewhat puzzling. The interval of .5 sec. between the manual response and the UCS should be effective for the conditioning of the eyelid response to the response-produced cues of the "mediating" response. The distribution of blink latencies to the CS is about what one would expect if this had occurred. Yet, response-produced-cue theory clearly predicts that CRs should occur when the mediating response is elicited by a different stimulus. The fact that this did not occur is a rather striking disconfirmation of conventional mediation theory. It may well be that response-produced cues play a far less important role than has commonly been thought, and are, in fact, relatively ineffective in the evocation of learned behavior. This, of course, would not deny the importance of kinesthesis in such things as the maintenance of posture and in learned and unlearned aspects of muscular coordination. To settle this question, there appears to be a need for further studies with overt mediating responses, involving appropriate manipulation and control of such variables as the temporal relations between the responses, the vigor and discriminability of mediating reactions, and the type of response employed, both in the roles of mediating and mediated responses.

REFERENCES

GREGG, L. W., & BROGDEN, W. J. The relation between duration and reaction time difference to fixed duration and response terminated stimuli. *J. comp. physiol. Psychol.*, 1950, **43**, 329–337.

GRICE, G. R., & DAVIS, J. D. Mediated stimulus equivalence and distinctiveness in human conditioning. *J. exp. Psychol.*, 1958, **55**, 565–571.

GRICE, G. R., & DAVIS, J. D. Effect of concurrent responses on the evocation and generalization of the conditioned eyeblink. *J. exp. Psychol.*, 1960, **59**, 391–395.

GRICE, G. R., SIMMONS, H. J., & HUNTER, J. J. Failures to obtain mediated generalization effects in eyelid conditioning. *J. exp. Psychol.*, 1963, **65**, 485–489.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

MILLER, J., & COLE, L. E. The influence of a "voluntary" reaction upon the development and the extinction of the conditioned eyelid reaction. *J. genet. Psychol.*, 1936, **48**, 405–440.

OST, J. W. P. Interstimulus interval and proportion of non-reinforced trials in differential eyelid conditioning. Unpublished doctoral dissertation, Duke University, 1960.

(Received October 11, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 347–352

SPATIAL STEREOTYPES OF FOUR DIMENSIONS OF PURE TONE[1]

S. A. MUDD[2]

Purdue University

4 dimensions of sound (frequency, intensity, duration, and direction—interaural intensity difference) were scaled for equal discriminability and tested for associative spatial stereotype among undergraduate Ss. The frequency dimension was found to have both a horizontal and a vertical stereotype, the latter being considerably stronger. Intensity and direction had, respectively, a vertical and a horizontal stereotype. The duration dimension had no spatial stereotype as defined in this set of experimental operations.

The experimental determination of the "natural potential" of a set of stimuli to evoke particular perceptual or perceptual-motor responses can be termed "testing for population stereotype." Testing for stereotype reveals basic habits or expectancies of response to particular stimuli. The operational measurement of population stereotype involves (Fitts & Deininger, 1954),

> determination of the relative frequency with which each permissible response is made [by naive *S*s] to each stimulus in a situation in which *E* gives no indication of what is considered an appropriate response [p. 484].

A knowledge of stereotyped responses to particular stimuli is particularly useful in the design of those situations which involve the transmission of information by means of various symbols. In relation to human engineering problems, the test-for-stereotype procedure has been used successfully in the design of pictorial road signs (Brainard, Campbell, & Elkin, 1961) and the development of a "picture language" to identify vehicle controls (Mudd & Karsh, 1961).

The present study was conducted as a step in the development of a pure tone auditory cuing system to be used in conjunction with an instrument monitoring problem. In such a visual monitoring system, coded auditory signals might be used to direct visual attention to the areas of an instrument panel that require immediate attention. Specifically, the experiment was designed to determine the extent to which four dimensions of pure tone (frequency, intensity, duration, and interaural intensity difference) had a "natural tendency" to be associated with the horizontal and vertical dimensions of space. That is, was a spatial meaning stereotype associated with the four sound dimensions investigated? If there proved to be reliable spatial associations for the auditory dimensions tested, this information could be used in the coding of auditory cues in a visual monitoring task.

Before testing for stereotype the four sound dimensions were scaled for equal discriminability (Mudd,

---

[1] This research was completed for partial fulfillment of the requirements for the PhD degree, June 1961, Purdue University under the direction of E. J. McCormick.

The overall research project was supported in turn by the Purdue Research Foundation and the National Science Foundation (1959–60, 1960–61, respectively). The writer is grateful to those responsible and to Max Steer and his staff of the Purdue Speech and Hearing Clinic for the use of their facilities and technical assistance.

[2] Now at Gettysburg College.

1961). The scaling sample involved a total of 42 undergraduate *S*s. The range scaled for each of the sound dimensions was as follows: intensity, 20 to 100 db. SPL; duration, 400 to 2,382 msec.; direction, +8 to −8 db. (interaural difference re *S*'s left ear); frequency, 312 to 8,389 cps. The results of the scaling operations are shown in Fig. 1. The reliabilities of the scales

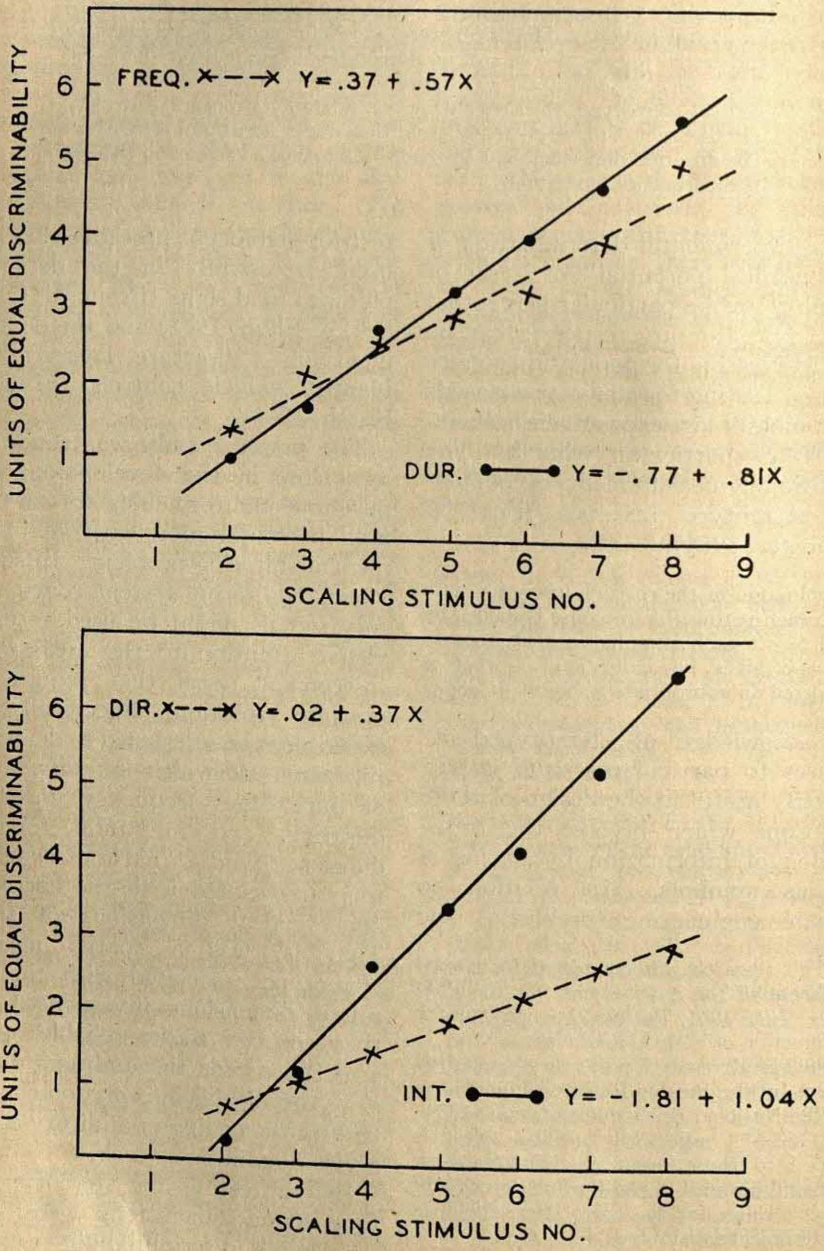


FIG. 1. Psychophysical functions.

as estimated by the absolute average discrepancy index (Torgerson, 1958, p. 241) were as follows: intensity, 0.5%; duration, 1.4%; direction, 3.4%; frequency, 1.8%. All sound dimensions scaled satisfactorily and the resultant equal-discriminability values were used for the systematic selection of comparable stimuli both within and across the four dimensions of sound to be tested for spatial stereotype.

METHOD

The situation used to define spatial stereotypes was as follows. The *S*s listened to two auditory stimuli which differed along just one dimension of sound (e.g., intensity). After listening to the stimulus pair, *S*s were asked to move a plug from a center reference hole in a pegpanel, which position "represented" the first (standard) stimulus, and plug the peg into another hole in the pegpanel to "represent" the second (comparison) stimulus. The spatial difference between the two pegholes was taken to "represent" the difference between the auditory stimuli being compared. Plotting was entirely arbitrary on the part of *S*s. The only specific "set" established by the experimental instructions was that *S*s should try to be consistent. The position of *S*s' plots was recorded in terms of *X* and of *Y* deviations from the zero origin (the spatial counterpart of the standard tone) of a coordinate system superimposed on the back of the pegpanel. Plot coordinates were averaged over three randomized plots by each *S* for each of six stimulus values along each of the four sound dimensions. The two sets of average measures (*X* and *Y* deviations) were subjected to simple analyses of variance in order to determine whether or not levels of sound stimuli gave rise to significant variance on either the horizontal or the vertical axis. The procedure had no direct precedent in the research literature for the spatial stereotype tested.

Stimulus values were selected so that the working range of each of the sound dimensions was divided into six equal sections balanced around the median stimulus value (in ED).[3] This procedure "normalized" the range used for each dimension by dividing each range by a factor of 6. Table 1 shows the stimulus values in ED units as well as the corresponding physical stimulus values. The standard reference tone was made up of the median stimulus values (in ED units) of all the auditory dimensions; it was repeated twice (two "on" periods of the duty cycle of an electronic switch). The standard reference tone was presented approximately 2 sec. before the variable comparison tone in each trial. The time lapse between trials varied from 5 to 8 sec. due to the variability inherent in manual programing.

TABLE 1

EQUALLY-DISCRIMINABLE STIMULI IN ED UNITS WITH CORRESPONDING PHYSICAL STIMULUS VALUES BY AUDITORY DIMENSION

	Stimulus No.						
	1	2	3	Standard	4	5	6
Frequency							
ED	.50	1.25	2.00	2.75	3.50	4.25	5.00
cps	222	402	748	1391	2574	4787	8861
Intensity							
ED	.75	1.75	2.75	3.75	4.75	5.75	6.75
db. SPL	35	44	54	64	73	83	92
Duration							
ED	.25	1.25	2.25	3.25	4.25	5.25	6.25
msec.	424	558	736	968	1276	1680	2215
Direction							
ED	.02	.76	1.50	1.87	2.24	2.98	3.72
db. (re left ear)	10	6	2	0	−2	−6	−10

Four semirandom programs were used in counterbalanced order across *S*s and conditions of sound dimension. The restriction imposed on completely random presentation of stimuli within conditions was such that no more than three stimuli in sequence should be from just the upper or just the lower part of the dimension being presented. The standard reference tone marked the upper/lower dichotomy. Each comparison tone within a given condition of sound dimension was presented three times. Each series of trials within given dimensions consisted of a total of 18 trials. The series for each dimension in turn was completed before going on to another dimension. The order of presentation of the dimensions was counterbalanced across *S*s.

---

[3] A minor exception was required in the case of the direction dimension. The apparatus used to vary interaural intensity differences did not permit the adjustments necessary for strict adherence to design. In that condition, Stimulus No. 3 and 4 are only one half as far apart on the stimulus range as the difference required by the design.

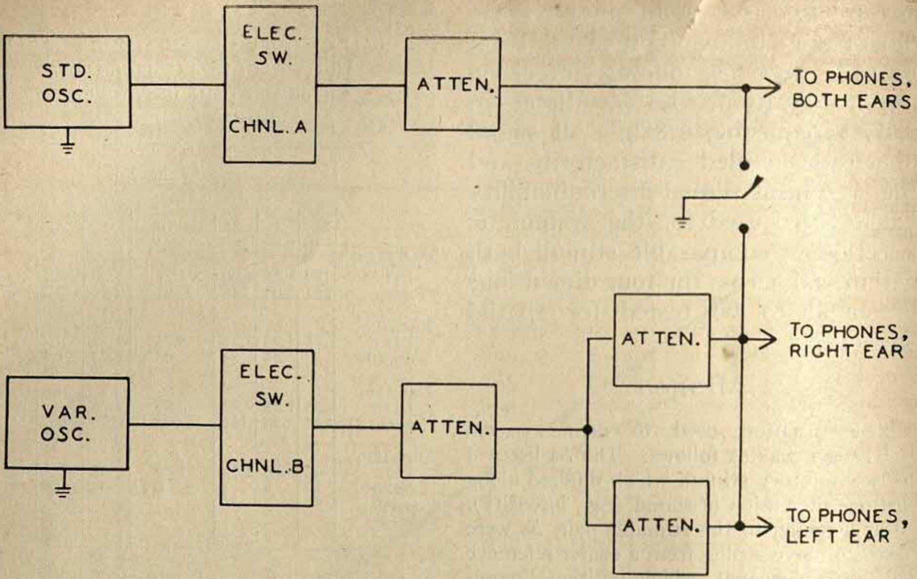


FIG. 2. Block diagram of circuit layout.

Apparatus.—A ⅛-in. sheet of Masonite consisting of a 19 × 19 matrix of pegholes which received a "banana-plug" made up the plotting panel. The holes were positioned symmetrically at 1-in. distances along both the vertical and horizontal dimension of the panel. Experimental sound stimuli were generated by means of primary generating equipment which consisted of two audio oscillators (Hewlett-Packard, Model 200CD), an electronic switch (Grason-Stadler, Model 829S-96), and three attenuators (Hewlett-Packard, Model 350-A). A block diagram of the circuits is presented in Fig. 2. All signal programing was done manually by *E*. The experimental stimuli were delivered across a multipole switch into a single set of binaural phones (Telephonics, Model T.D.H. 39-300Z).

TABLE 2

MEAN COORDINATE VALUES IN INCHES FOR EXPERIMENTAL STIMULI BY DIMENSION OF SOUND

	Stimulus No.					
	1	2	3	4	5	6
Frequency						
X	−2.65	−1.02	−.67	1.42	2.29	1.77
Y	−5.12	−2.68	−1.66	3.67	5.05	5.76
Intensity						
X	−.91	−1.10	−.56	.35	1.41	1.69
Y	−3.36	−2.71	−1.97	.94	1.74	3.36
Duration						
X	−1.39	−.42	.39	.50	1.00	1.64
Y	−.74	.36	.32	.46	.47	.32
Direction						
X	−4.38	−3.32	−.79	2.22	3.61	4.57
Y	.25	.20	−.15	−.26	.07	.18

Subjects.—A total of 31 *S*s served in the study (13 males, 18 females); all *S*s were Purdue undergraduates. Each *S* plotted each of the four sound dimensions evaluated. The data for one *S* for the direction dimension were discarded (a lead to one of *S*'s phones had opened during the experimental session). It is possible that the sample used was biased to the extent that the typical Purdue student has had more experience with the use of point representation on an *X*, *Y* coordinate system as compared to the general population.

Procedure.—The *S*s were seated in the listening position. A copy of general instructions was given them. The *E* read the instructions aloud as *S* followed on his copy. The general instructions contained orientation information. A second set of specific instructions was then presented in the same manner. The specific verbal instructions were supplemented with cartoon illustrations to show *S*s "how" to plot (*not* where). The *S*s were then fitted with headphones. The first experimental series was run with no further communication between *S* and *E*. At the completion of the first series, the range of the second series was introduced and that series run. The procedure was followed until *S*s had plotted each of the four sound dimensions.

Experimental design.—The plots making up the raw data were in the form of *X*, *Y* coordinates corresponding to the horizontal and vertical dimensions of the pegpanel with the origin at the center reference position; each unit displacement along each axis equaled 1 in. Data yielded by an *S*'s three plots of each stimulus value were reduced to an average vertical displacement and an average horizontal displacement. The average values across *S*s are shown in Table 2. Eight simple analyses of variance were used to evaluate the data, one analysis of the vertical component of the plots and one analysis of the horizontal values for each of the four sound dimensions. A summary of these analyses is provided in Table 3.

RESULTS AND DISCUSSION

A significant *F* ratio found in conjunction with either, or both, spatial coordinate measures was interpreted as indicating that a spatial stereotype existed for the sound dimension concerned. That is, to the extent that levels of a given sound dimension gave rise to significant variance among the mean spatial plots along either the vertical or the horizontal axis of the pegpanel, then it seemed reasonable to consider that a spatial stereotype existed in the population for that sound dimension.

In terms of that criterion the duration dimension may be said to have had no spatial stereotype since the *F* ratio for both the vertical and the horizontal measure did not exceed a significant level, *F* (5, 180) = .35 and 2.25, respectively. Direction may be said to have had a strong horizontal stereotype since the mean plot variance along that spatial axis far exceeded the .01 level of probability, *F* (5, 174) = 60.03, and the corresponding *F* value for the vertical dimension was not significant, *F* (5, 174) = .38. The intensity dimension yielded a significant *F* ratio for the vertical axis, *F* (5, 180) = 8.83, and an insignificant value for the horizontal, *F* (5, 180) = 1.85. These values for the intensity dimension were interpreted as being indicative of a vertical stereotype.

TABLE 3

ANALYSES OF VARIANCE FOR PLOT DEVIATIONS ON BOTH THE *X* AND *Y* AXES FOR EACH OF FOUR SOUND DIMENSIONS

Source	*df*	*MS*	*F*
Horizontal (*X*) deviations			
Frequency	5	124.15	7.20*
Error	180	17.24	
Intensity	5	44.60	1.85
Error	180	24.13	
Duration	5	37.25	2.22
Error	180	16.81	
Direction	5	415.39	60.03**
Error	174	6.92	
Vertical (*Y*) deviations			
Frequency	5	665.25	38.72**
Error	180	17.18	
Intensity	5	229.75	8.83*
Error	180	26.01	
Duration	5	6.67	.35
Error	180	19.27	
Direction	5	1.28	.38
Error	174	3.39	

* $p < .05$.
** $p < .01$.

The final sound dimension tested for spatial stereotypy, frequency, required special consideration in that both *F* ratios were significant. Apparently frequency was associated with the horizontal dimension of space as indicated by an *F* (5, 180) = 7.20, $p = .05$. At the same time frequency seemed to have had an even stronger affinity for the vertical, the *F* (5, 180) for that measure = 38.72, $p < .01$. From a theoretical point of view this latter finding seems not to introduce any particular difficulty. There is no reason why any set of stimuli might not have more than one consistent associative response. In fact a logical extension of this type of study would involve testing the same sound dimensions for depth associations, a third pertinent dimension of space not considered in this study. It seems reasonable to expect that some of these sound characteristics would also connote spatial depth.

From a practical standpoint it would be important to recognize that the spatial stereotype evoked by the frequency characteristic was essentially multidimensional. That is, frequency tended to evoke both "vertical" and "horizontal" meaning, although the vertical connotation, on the basis of a considerably larger *F* ratio, was much stronger than the horizontal. In a practical situation, such as the development of a set of pure tone stimuli to be used to signal spatial location, it would seem more reasonable to assign frequency as a vertical cue since the spatial stereotype for the vertical as defined in this study was much stronger than for the horizontal.

One further point of discussion requires consideration here; this concerns the adequacy of the procedure used as a measure of stereotype. An essential part of the procedure requires that *S*s be completely free to respond as they will to the presented stimuli—that *E* provide no indication of what is considered to be an appropriate response. To the extent that *E* by means of the particular instructions does provide *S*s with a response set, there is the possibility that the stereotype response set could be masked, or biased. Two precautions were taken in this study to forestall such a biasing effect. First, the verbal instructions emphasized that *S*s could and should plot anywhere on the panel. Second, the cartoon illustrations used to facilitate understanding of the task were designed to avoid, or at least to hold constant, any such tendency to establish a response set. Of four cartoon figures involved, only one depicted *S* holding the plotting peg, and the position of the peg in that cartoon was deliberately put midway between the *X* and the *Y* axis in the third quadrant. It was believed that if there were any biasing influence in the illustration it would operate across both the spatial dimensions, and in effect be controlled.

Generally it is felt that the above procedures constituted an adequate test for spatial stereotype and that a similar procedure might profitably be used to determine the depth stereotype for these same and other stimulus dimensions, both auditory and otherwise. The procedure provides an empirical basis for the assignment of vertical or horizontal cuing value to the tonal characteristics tested.

REFERENCES

BRAINARD, R. W., CAMPBELL, R. J., & ELKIN, E. H. Design and interpretability of road signs. *J. appl. Psychol.*, 1961, **45**, 130–136.

FITTS, P. M., & DEININGER, R. L. S-R compatibility: Correspondence among paired elements within stimulus and response codes. *J. exp. Psychol.*, 1954, **48**, 483–492.

MUDD, S. A. The scaling and experimental investigation of four dimensions of pure tone and their use in an audio-visual monitoring problem. Unpublished doctoral dissertation, Purdue University, 1961.

MUDD, S. A., & KARSH, R. Design of a picture language to identify vehicle controls: I. General method. II. Investigation of population stereotypes. *USA Ordn. Hum. Engng. Lab. tech. Memo.*, 1961, No. 22–61.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

(Received October 15, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 353–362

SERIAL LEARNING: POSITION LEARNING AND SEQUENTIAL ASSOCIATIONS

SHELDON M. EBENHOLTZ[1]

Connecticut College

A serial list may be mastered by learning the location of the items relative to the beginning and end of the list, by forming sequential associations between successive terms, or on the basis of both processes. 2 experiments yielded evidence for the role of position learning as opposed to sequential associations. 1 experiment compared the usual method of serial learning with one in which S began each trial at a different point in the sequence. The latter procedure produced significantly greater difficulty even though sequence and number of associations were identical in both conditions. The 2nd experiment demonstrated faster learning for items on a 2nd list which occupied identical positions on both original and transfer lists than for items appearing at noncorresponding positions. In addition, a test of associative interference under conditions of serial learning yielded negative results thereby failing to support the sequential association hypothesis.

Traditionally, investigations of serial learning have shared the assumption that associations are formed between successive members of the series. In addition, extensive use has been made of the assumption of the formation, loss, and subsequent recovery of S-R bonds (e.g., Underwood & Postman, 1960) in dealing with the retention of material acquired under conditions of serial learning. There appears, however, to be little direct, unambiguous evidence for the assumption that the establishment of consecutive associations is fundamental to the acquisition of a serial list.

The problem becomes more pronounced in light of an alternate conception of serial learning, which would suggest that the task is described more adequately in terms of a position learning process: that sequential associations are essentially by-products of the major task which is to learn the relative locations of the various items. According to this view, a serial list may be mastered by learning the items and their locations relative to the beginning and end of the sequence. The process which is advanced here may be likened to one in which a set of items (i.e., nonsense syllables, words, objects, etc.) is distributed within some spatial context. If the items are arranged so that the set as a whole contains identifiable end points, as for example in a straight line, then the task of learning the items and their locations may be accomplished without ever establishing specific associations between the terms comprising the set. It is possible that a similar process occurs in the case of serial learning. Learning of the positions of items which are presented in a temporal order may be considered to be strictly analogous to the learning of position within a spatial arrangement. Since knowledge of the positions of the component items entails knowledge of the sequence, it is apparent that position learning could lead to the same outcome (mastery of a serial list) as that which

---

[1] The author is grateful to Irvin Rock and John Ceraso for their many helpful suggestions. Thanks are extended to Roy Malpass for his aid in obtaining *S*s for the various experiments.

would be obtained were specific sequential associations established.

Recent years have provided relatively few instances (Asch, Hay, & Diamond, 1960; Schulz, 1955; Wishner, Shipley, & Hurvich, 1957; Woodworth & Schlosberg, 1954, p. 711) in which it has been suggested that position learning may play an important role in the serial learning process. In the present paper evidence for the role of position learning in serial learning will be presented. At the same time the generally held view that the formation of sequential associations more typically describes the serial learning process will be evaluated.

In the first experiment two conditions of serial learning were compared. Two groups of *S*s were given the task of learning an *identical sequence*, however, in one case the location of the items with respect to the beginning, middle, end, etc. of a list was varied from trial to trial. In the other case the location of items was held constant. If the standard method of serial learning entails primarily the formation of sequential associations, few differences should result between the two procedures. However, large differences should result if position learning plays an important role.

In the second experiment the attempt was made to establish further support for the role of position in serial learning. The assumption of position learning implies that in the learning of a second serial list containing items repeated from a first learned list, those items which are repeated at coordinate serial positions should be learned earlier than transferred items which occur at disparate serial positions. In addition, Exp. II was designed to investigate the operation of specific associative interference (AI) in the context of serial learning. The hypothesis tested was that if associations are formed between successive members of the series, then a transfer task requiring the substitution of new for old responses should lead to a performance decrement.

EXPERIMENT I: SERIAL LEARNING WITH VARIABLE SERIAL POSITION

Method

Procedure.—The procedure enabled *E* to compare the learning of an *identical* sequence of syllables under two different conditions. Control *S*s learned by the usual method of serial anticipation. Experimental *S*s also learned by the method of anticipation, however, the particular syllable which started the list on a given trial was varied from trial to trial.

The sequence of syllables as they appeared for the Experimental *S*s on Trials 1, 2, and 10 may be represented as follows:

Trial 1. A-B-C-D-E-F-G-H-I-J-A-B-C
Trial 2. D-E-F-G-H-I-J-A-B-C-D-E-F
. . . .
Trial 10. H-I-J-A-B-C-D-E-F-G-H-I-J

Each letter stands for a nonsense syllable. Note that although the syllables change their positions with respect to the beginning and end of each trial, the stimulus response relation between an item and its immediate neighbor is preserved. The trial to trial sequence maintains the intratrial order as well. On each trial the first and last three syllables are duplicated. The purpose of this arrangement was to allow *S* to orient himself on each trial to the changing context. The first three syllables could get *S* started on the proper sequence.

There were 12 *S*s in the Experimental group (Group E). One *S* started on each of the 10 trial orders; 2 additional *S*s initiated learning on Trial Orders 1 and 2. All *S*s learned, by spelling, to a criterion of 12 correct responses on a single trial. Syllables were presented at a 2-sec. rate with a 14-sec. intertrial interval. The first syllable on each trial served as the cue to initiate anticipation.

Each of the 10 *S*s in the Control group (Group C) was presented with one of the 10 trial orders represented above. Thus both groups learned equal sized lists of identical content and to the same criterion. For Group C both the S-R relations as well as the location of syllables were constant from trial to trial.

With the exception of the first two *S*s who performed under the Experimental condition *S*s were alternately assigned to Group E and Group C in order of their appearance at the laboratory. All were naive to the purpose of the study and had no previous experience with verbal learning experiments.

Materials.—The lists were made up of 10 different syllables taken from the Glaze (1928) list of 80–87% association values. Each of the five vowels occurred twice; no syllable shared a first or third consonant with a first or third consonant of any other syllable.

Results

Table 1 shows the mean trials to criterion, mean errors per trial, and mean intrusions per trial. We will take these up in order.

Trials to criterion.—Group E required a mean of 35.9 ($SD = 17.1$) trials to achieve the criterion. Group C required a mean of 22.7 ($SD=9.4$) trials to reach the same level of mastery. The difference is significant, $t(20) = 2.19$, $p < .02$. Since there was some tendency towards heterogeneity of variance, the Mann-Whitney *U* test was applied to the data. The conclusion was corroborated; $p < .05$.

Figure 1 shows the number of trials required to reach successive levels of performance for the two groups. The superiority of Group C is evident at all levels of mastery.

TABLE 1

MEAN TRIALS TO CRITERION, MEAN ERRORS PER TRIAL, AND MEAN INTRUSIONS PER TRIAL

	Control		Experimental	
	M	*SD*	*M*	*SD*
Mean trials to criterion	22.7	9.4	35.9	17.1
Mean total errors per trial	5.5	.8	6.8	1.1
Mean intrusions per trial	1.1	.3	1.1	.8

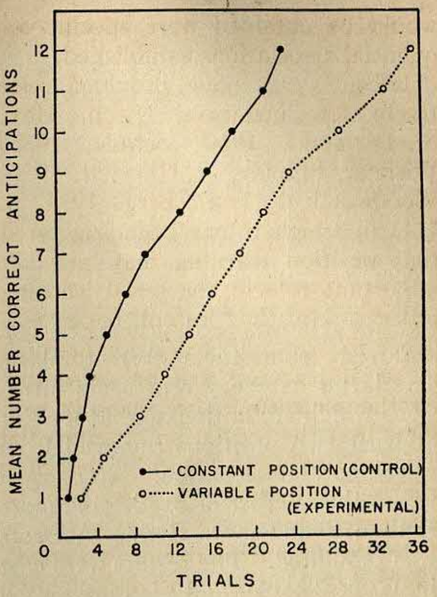


FIG. 1. Mean trials to reach successive criteria for constant and variable position groups.

Total errors.—Table 1 shows that the mean errors per trial (including omissions) is greater for Group E (6.8) than for Group C (5.5). On the basis of the *U* statistic the difference is significant ($p < .02$). We conclude that serial learning with variable position (Experimental group) produces more errors per trial than is the case under standard serial learning conditions.

Intralist intrusions.—The mean number of intrusions per trial produced by the two groups was identical, each group averaging 1.1 intrusions. However, the distribution of intrusions according to degree of remoteness from their correct place, did differ in the two groups. The intrusions were divided into two categories: Adjacent (Type A) and Remote (Type R). The former refers to those items that are erroneously anticipated

TABLE 2

NUMBER OF *S*s YIELDING MORE REMOTE (R) THAN ADJACENT (A) INTRUSIONS AND VICE VERSA

	Group C	Group E
A > R	8	4
R > A	2	8
N	10	12

at a point no more than one position removed from their correct place. The latter category represents intrusions which are anticipated at a point at least two positions removed from their correct location. For each group the number of *S*s yielding more A than R and more R than A errors was tabulated. The results were cast in a 2 × 2 table, and are represented in Table 2.

Table 2 shows that for Group C, 8 out of 10 *S*s make more Type A intrusion errors than Type R; for Group E, the opposite seems to be the case, 8 out of 12 *S*s yielding more Type R than Type A errors. The difference between the two distributions was significant (Fisher exact probability test; $p < .05$).

Discussion

The results indicate that serial learning is inhibited when the location of the items is changed from trial to trial. Since the number and kinds of associations were identical in the two groups, the obtained differences point up the contribution of constant temporal position in serial learning.

It may be argued that the procedure of constantly changing the starting point, in the experimental condition, may have disturbed or otherwise inhibited learning by interfering with the associative structure of the list. This is a particularly plausible possibility if one views the serial list as a complex of remote and adjacent connections. According to this view, changes in the starting point could serve to inhibit the arousal of preceding associations, a condition necessary for correct anticipation. The obtained differences in rate of learning would then be compatible with an associative interpretation as well as with the view presently suggested. The lists were constructed with this possibility in mind and they, therefore, contained three syllables which were repeated at the beginning and end of each trial. This allowed *S* some time in which to "get started" on the sequence. It is conceivable, therefore, that on a given trial, *S*s of Group E could anticipate as many syllables as those of Group C, as far as the early stages of list acquisition are concerned. That superior performance of Group C is manifest at all stages of list mastery (see Fig. 1) reduces the plausibility of this possibility.

A somewhat more detailed analysis of the data was undertaken as a further attempt to investigate this argument. It was assumed that a reasonable interpretation of the associative view was that anticipation of the last six syllables of the experimental list would derive full benefit of the remote associations of the first half of the list. Thus the two groups should show minimal differences when compared on the latter half of each list. To test this hypothesis the mean number of trials to anticipate correctly the last six syllables in sequence, was calculated. The results were 27.4 ($SD = 11.6$) for Group E and 15.8 ($SD = 13.8$) for Group C. The difference was significant, $t(20) = 1.98$, $p < .05$. The present results offer no support for an associative interpretation of the main findings.

The fact that the groups do not differ in mean intrusion errors per trial is compatible with the conventional view of serial learning. This is especially clear when one identifies intrusions with remote associations; since the two conditions may be regarded as equivalent with respect to the adjacent as well as remote association structure within the lists, frequency of misplacements should likewise be equivalent. We noted, however, that the two groups differed in the distribution of Type R and Type A errors. The Type A errors were more frequent for Group C while Type R errors were more frequent for Group E. This finding suggests that Group C responses were selective with respect to serial position. For Group E the use of serial position as a basis of responding, was largely precluded by the changing orders of presentation. However, many of these *S*s did report that one or sometimes two syllables gradually took on the function of a "subjective" beginning to the sequence. This suggests that to some extent position learning could have occurred under the Experimental as well as Control conditions.

While these results do not warrant the conclusion that serial learning is based entirely upon the learning of position, they likewise do not support the contention that mastery of a serial list entails the formation of sequential associations exclusively. We shall return to this point later. For the present we conclude that elimination of position as a basis for responding, produces large and significant decrements in the rate of learning a serial list.

EXPERIMENT II: POSITIVE TRANSFER AND ASSOCIATIVE INTERFERENCE IN SERIAL LEARNING

Method

Procedure.—Thirty-six Union College students took part. These were distributed into three groups according to their order of appearance at the laboratory. Each *S* learned two serial lists successively. The following paradigm is representative of Group 1 conditions.

List 1—A B C D E F G H I J A
List 2—K_1B K_2D K_3F K_4H K_5J K_1

In this and the following illustrations, letters of the alphabet are used to represent nonsense syllables. Note that every other syllable on List 2 was repeated from List 1 and occupied identical positions on both lists, whereas alternate syllables were new. The latter are represented by the letter K with subscripts indicating five different new items.

Group 2 *S*s were given lists constructed according to the following paradigm.

List 1—A B C D E F G H I J A
List 2—F K_4H K_5J K_1B K_2D K_3F

As was the case with Group 1, every other syllable of List 1 was repeated on List 2. However, unlike Group 1 the repeated syllables occurred a minimum of four positions removed from their original location on List 1.

An important property of the first list was that it began and ended with the same syllable. The items, therefore, constituted a closed "chain" of associations, i.e., a list in which each item including the last was associated with a succeeding term. For example, on List 1 Item J was directly associated with A and, therefore, remotely associated with B etc. By virtue of this property both groups were equated for the role of remote associations in the learning of repeated items in List 2. That is, since every other item on List 1 was repeated on List 2, the repeated terms constituted a sequence of one degree of remoteness. These sequences may be represented as B-D-F-H-J-B etc. and F-H-J-B-D-F etc., for Groups 1 and 2, respectively. For both groups, therefore, repeated items may be assumed to receive the benefit of remote associations established in the course of learning List 1.

Both groups as a whole learned *identical* first and second lists, respectively, although the specific List 1-List 2 pairings differed.

Half of each group learned repeated items which occupied odd numbered positions on List 2. The remaining *S*s learned lists with repeated items at even numbered positions. Learning was by the method of anticipation (spelling), to a criterion of one errorless trial. Six different list orders were employed.

The net results of the above procedure were to equate the groups for (*a*) specific content and (*b*) the effects of remote associations on the repeated items of List 2. The major difference in the treatment of the two groups was that the repeated items of Group 2 (List 2) occurred *out* of position with respect to their location on List 1, whereas repeated items of Group 1 maintained identical positions on both lists.

If it is correct that *S* learns that the various items belong at specific locations in the time series, then if the same items appear in identical locations on a second serial list, (e.g., Group 1) they should manifest positive transfer. Since repeated items of Group 2 occurred at different positions in the initial and transfer lists, positive effects of position learning were not expected to occur for this group.

TABLE 3

Performance on List 1: Mean Trials to Criterion and Mean Errors per Trial

Group	Trials to Criterion		Errors Per Trial	
	M	*SD*	*M*	*SD*
1	26.4	14.8	5.4	1.0
2	28.2	10.2	5.6	1.2
3	31.8	13.7	5.8	1.4

On the assumption that sequential associations are established in the course of learning a serial list, it follows that the learning of new items (List 2) should exhibit associative interference. This holds equally for Group 1 and Group 2. An S-R view likewise yields no differential prediction in the case of repeated items viz., both groups should perform equally well.

In order to permit an estimate of the absolute level of transfer, a control group was employed. These *S*s learned two serial lists, the second of which was identical with the second lists of Groups 1 and 2. Their first list was, however, different, so that no items were repeated between List 1 and List 2. Since all groups learned identical second lists it was possible to make separate comparisons based upon the learning of "new" items and in terms of items designated "repeated." Transfer effects may, thus be represented by differences between Experimental (Groups 1 and 2) and Control groups (Group 3).

Materials.—All lists were composed of 11 nonsense syllables (including the cue syllable). For Groups 1 and 2, List 1 was composed of syllables which had an average association value at 33.5% (Glaze, 1928). The initial list of Group 3 averaged 35.3%. Six *S*s in each of the three groups learned second lists with an average association value of 46.9%. The average association value of List 2 for the remaining half of each group was 34.8%. The difference in association value was due to the fact that for half the *S*s, List 2 contained repeated List 1 items which were *not* repeated for other *S*s. This in turn resulted from the fact that half of the *S*s repeated odd positioned List 1 items whereas the rest of the *S*s repeated even positioned List 1 items.

A 3-sec. presentation rate was employed with a 27-sec. intertrial interval. List 2 was presented within 1 min. after completion of List 1.

Results

Original list.—Table 3 presents mean trials to criterion and mean errors per trial for all three groups on List 1.

Separate analyses of variance for trials to criterion and errors per trial yielded nonsignificant F ratios; F $(2, 33) = .48$ and .44, respectively. Groups 1 and 2 learned identical lists and since Group 3 *S*s learned lists of comparable association value, we conclude that the three groups did not differ in learning ability.

Transfer list.—In the following analysis, new items of all three groups were compared directly with each other, as were repeated items. Group 3 (Control) learned no repeated items since the entire list was new to these *S*s. However, all three groups learned identical transfer lists. This allowed for the identification in Group 3 of those items which had been repeated in Groups 1 and 2. Similarly, the designation "new items" when applied to Group 3 simply refers to those syllables of Group 3 which were new for Groups 1 and 2.

1. Trial of first correct response. Table 4 represents the average trial on which new and repeated items were first given correctly. Separate analyses of variance for new and repeated items yielded $F\,(2, 33) = 1.51$ and 9.11, respectively. The former value was not significant, $(p > .05)$. In

TABLE 4

Mean Trial of First Correct Response for New and Repeated Items Separately

Group	New		Repeated	
	M	*SD*	*M*	*SD*
1	5.0	1.8	3.8	1.5
2	5.7	2.3	6.6	1.9
3	6.7	2.6	7.1	2.4

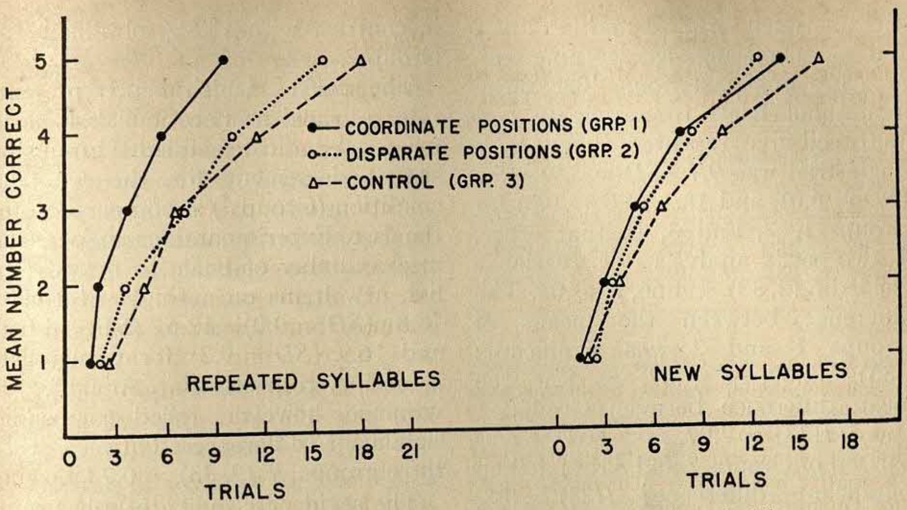


FIG. 2. Mean number of correct anticipations as a function of trials; repeated and new syllables separately.

the case of repeated items the *F* ratio of 9.11 was highly significant ($p < .01$). Evaluation of the differences between the means of Groups 1, 2, and 3 (repeated items) was based upon the within *MS*. All reported p values were drawn from a one-tailed test of significance. The difference between the means of Groups 1 and 2 of 2.8 was significant; $t(33) = 3.37$, $p < .002$, as was the difference of 3.3 between Groups 1 and 3; $t(33) = 3.97$, $p < .0002$. The difference of .5 between Groups 2 and 3, however, did not reach significance; $t(33) = .60$, $p > .05$.

Analysis of new items indicated no significant differences in the trial of the first correct response. We conclude, therefore, that the measure employed yielded no evidence of associative interference under the present conditions of serial learning.

The learning of repeated items clearly demonstrates positive transfer only for those syllables that are repeated *in position* (Group 1). Syllables repeated at positions which differ from their original List 1 location (Group 2) are learned no earlier than items which are learned for the first time (Group 3). It follows, therefore, that under the present conditions of serial learning, positive transfer is a function of the maintenance of identical serial positions on original and transfer lists.

2. Rate of mastery of new and repeated syllables. It is conceivable that the effects demonstrated above, may be transitory. It is also possible that the sensitivity of the measure as an indicator of positive and negative transfer may vary as a function of the criterion selected (Weitz, 1961). It is, therefore, desirable to compare the groups on the basis of a measure of learning which represents the entire course of acquisition.

Figure 2 represents successive criteria at which repeated and new syllables, respectively, were mastered. Each curve is composed of five points since *S*s learned a maximum of five new and five repeated syllables only.

In the case of repeated syllables, the curves of Fig. 2 show a rapid rate of learning for Group 1 *S*s; the differences between Group 1 and Groups 2 and 3 increasing as learning approaches final criterion. The mean number of trials to anticipate correctly all five repeated items on a single trial was 9.7 (*SD* = 3.9), 15.6 (*SD* = 6.0), and 18.1 (*SD* = 9.0) for Groups 1, 2, and 3, in that order. A two-part analysis of variance yielded $F\,(2, 33) = 4.66, p < .05$. The difference between the means of Groups 1 and 2 was significant; $t(33) = 2.09$, $p < .02$, as was the mean difference between Groups 1 and 3; $t(33) = 2.97, p < .002$. Comparison of Groups 2 and 3 revealed no significant difference; $t(33) = .88$, $p > .1$. Since the data indicated heterogeneity of variance, the Kruskal-Wallis one-way analysis of variance (Siegel, 1956) was applied. The identical pattern of significance as reported above, was obtained.

A measure of the magnitude of positive transfer in Group 1 may be given by expressing the difference between Groups 1 and 3 (Control) as a percentage of Group 3. This yields 46.4%, a substantial amount in contrast with 13.8% obtained for Group 2.

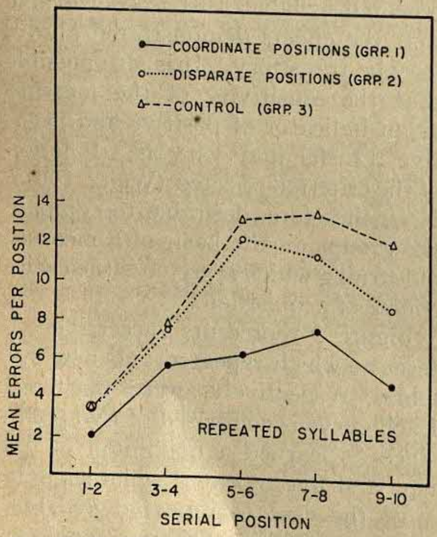


FIG. 3. Serial position curves for repeated items of Groups 1, 2, and 3, respectively.

The curves which represent the rate at which new syllables are mastered indicate a slight but consistent decrement for the Control condition (Group 3) as compared with the two Experimental groups. The mean number of trials to master all five new items on a single trial was 13.8 (*SD* = 9.0), 12.6 (*SD* = 7.6), and 16.9 (*SD* = 9.2) for Groups 1, 2, and 3, respectively. Analysis of variance, however, failed to reveal significant differences between the three groups; $F\,(2, 33) = 0.73$.

The evidence indicates that repeated syllables are mastered more rapidly when they occur at coordinate serial positions (Group 1) than at disparate serial positions (Group 2); the latter are learned at the same rate as control syllables. The hypothesis of a position learning process is again confirmed. The absence of a difference between the three groups in the rate of learning new words indicates no effect of associative interference.

3. Serial position effects. Since high positive transfer of repeated items was apparently mediated by serial position, it is in order to inquire of the extent of such transfer across all positions. If the effect is limited to one or two positions only, the implications for a position learning interpretation of serial learning would be severely restricted. The following analysis was performed in order to assess this possibility.

The mean number of errors to criterion (i.e., to learn the entire list) for each *S* at each position in List 2, was tallied. The data for repeated syllables were recorded for Groups 1, 2, and 3, respectively. Figure 3 represents serial position curves with successive pairs of positions combined in order to increase reliability

(12 *S*s in each group contributed scores to each of the five points on the graph, 6 to each even numbered position and 6 to the remaining odd numbered positions).

The curves of Fig. 3 indicate high positive transfer for Group 1 *S*s. Serial position identity is effective at all positions in mediating such transfer, although the maximum gain apparently occurs at central and terminal positions.

Table 5 represents mean errors in learning the entire list, for new and repeated syllables, respectively.

Analysis of variance for new words yielded $F\,(2, 33) = 1.22$ which was not significant ($p > .05$). For repeated items $F\,(2, 33) = 3.23$ which was significant ($p < .05$). Individual *t* tests based upon the within *MS* for repeated syllables yielded the following: the mean difference of 16.9 between Groups 1 and 2 was significant, $t(33) = 1.76$, $p < .05$; comparison of Groups 1 and 3 showed a mean difference of 23.8 errors; $t(33) = 2.47$, $p < .01$. The difference of 6.9 errors between the means of Groups 2 and 3 was not significant; $t(33) = .71$, $p > .2$.

These results are quite consistent with the analyses dealing with the trial of first correct response and the rate at which successive criteria are attained.

TABLE 5

MEAN ERRORS TO CRITERION FOR NEW AND REPEATED ITEMS SEPARATELY

Group	New		Repeated	
	M	*SD*	*M*	*SD*
1	31.2	20.3	26.3	20.5
2	38.4	21.4	43.2	20.6
3	46.2	25.3	50.1	25.9

Discussion

The data suggest the following conclusions.

Positive transfer.—Repeated syllables which occurred at coordinate positions on both lists (Group 1) were learned significantly more rapidly than repeated items which occurred at disparate serial positions (Group 2). Since the groups differed in no way except in regard to the relative locations of repeated items, this result is taken to imply that the rapid learning manifest by Group 1 was mediated by a position learning process.

The rate at which repeated items of Group 2 were learned did not differ significantly from the rate of learning comparable items in the Control group. The absence of a difference, despite the fact that Group 2 had the benefit of both item familiarity and remote associations suggests that the transfer of items to disparate serial positions may in fact have *impeded* performance on the transfer task. Additional evidence, however, is needed on this point.

Group 1 learned repeated items significantly more rapidly than the Control group. This finding established that positive transfer did in fact occur (a conclusion which did not follow from comparison of Groups 1 and 2 alone). However, it is not known to what extent the absence of familiar syllables in the Control condition contributed to this outcome.

Negative transfer (associative interference).—Analyses of new items suggest that either associative interference does not occur in serial learning, or that present conditions were insensitive to its effects. However, the lack of evidence for the traditional S-R view of serial learning is consonant with the findings of Young (1961).

No decision is yet available as to the extent to which sequential association and/or position learning characterize serial learning. Since these alternatives may not be regarded as mutually exclusive, and since present evidence does not indicate the relative contribution of each process, traditional interpretations of serial learning phenomena must be regarded with caution.

REFERENCES

Asch, S. E., Hay, J., & Diamond, R. M. Perceptual organization in serial rote-learning. *Amer. J. Psychol.*, 1960, **73**, 177–198.

Glaze, J. A. The association values of nonsense syllables. *J. genet. Psychol.*, 1928, **25**, 255–269.

Schulz, R. W. Generalization of serial position in rote serial learning. *J. exp. Psychol.*, 1955, **49**, 267–272.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Underwood, B. J., & Postman, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

Weitz, J. Criteria for criteria. *Amer. Psychologist*, 1961, **16**, 228–231.

Wishner, J., Shipley, T. E., & Hurvich, M. S. The serial position curve as a function of organization. *Amer. J. Psychol.*, 1957, **70**, 258–262.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

Young, R. K. The stimulus in serial verbal learning. *Amer. J. Psychol.*, 1961, **74**, 517–528.

(Received October 26, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 363–370

SOME SOURCES OF ARTIFACT IN STUDIES OF THE TACHISTOSCOPIC PERCEPTION OF WORDS[1]

JAN PIERCE[2]

University of Illinois

A methodological study of tachistoscopic word thresholds. Stimuli were 5-letter words varying in frequency; Ss were male undergraduates. Experimental groups varied: (a) stimulus presentation—random or repeated, (b) size of stimulus population—list or no list, (c) stimuli per exposure—1 or 2, (d) task—identification or discrimination, and (e) response—forced or free. Frequency affected identification thresholds only when Ss had no word list. Identification of the first letter of the words showed no frequency effects. Identification thresholds were lower than discrimination thresholds. The forced-response method lowered identification thresholds for Ss with word lists. The relevance of the findings for studies of perceptual defense, vigilance, and subliminal perception was pointed out.

Starting with the earliest of the perceptual defense and vigilance studies (Bruner & Postman, 1947; McGinnies, 1949; Postman, Bruner, & McGinnies, 1948), critics have offered alternative explanations of the findings, generally based on more pedestrian threshold determinants such as stimulus frequency and set (e.g., Howes & Solomon, 1950). As the studies themselves have become more elegant and far-reaching (Cowen & Beier, 1954; Wiener, 1955), so have the critics' explanations of them (Eriksen, 1958; Goldiamond, 1958). Many of the original proponents of perceptual defense and vigilance and of such related phenomena as subliminal perception have been moved to reverse their position.

However, despite increasing focus on the importance of rigor and precision in determining thresholds, studies of defense, vigilance, and subliminal perception have continued to appear which fail to define operationally such terms as "level of conscious awareness" (Blum, 1954), or which define thresholds at accuracy levels far in excess of chance expectancy (Pustell, 1957).

In addition, there has been growing doubt about routinely accepted threshold determinants themselves. Eriksen and Browne (1956) pointed out the importance of differential response strength in studies of threshold and suggested that such studies were dealing primarily with guessing behavior rather than with perception. Along these lines, Goldiamond and Hawkins (1958) have cast doubt upon the perceptual nature of the often-invoked relationship between word frequency and word thresholds. They concluded that the frequency-threshold relationship is a response effect rather than a perceptual one.

Another question is that of the generality of threshold parameters. Blackwell (1953), for example, has shown that the forced-response method, where *S* must say in which of several intervals he thinks a stimulus occurred, produces lower brightness detection thresholds than does the free-response method, where *S* is free to say either "yes" or "no" for each interval. As typically applied to form identification thresholds, however, the forced- and the free-response methods differ merely in whether or not *S* is required to make a response for each stimulus presentation; that is, whether or not he must guess.

---

[1] This study is based on a doctoral dissertation submitted to the Department of Psychology, University of Illinois, and was partially supported by Research Grant M-1206 from the National Institute of Mental Health, Public Health Service. The encouragement and suggestions of Charles W. Eriksen are gratefully acknowledged.

[2] Now at University of Iowa.

Similarly, the relationship between form identification and form discrimination thresholds is often crucial to subliminal perception experiments, a number of which have been criticized (Eriksen, 1958) on the grounds that discrimination among stimuli may well be possible at levels below that for correct identification (Klein, 1959; Taylor, 1953).

The present experiment is an attempt to answer some of the above questions and to explore a number of sources of artifact in the study of tachistoscopic word thresholds. Specifically, it was anticipated that: (*a*) differences in threshold between high and low frequency words could be shown with a discrimination (same-different) method which precludes response artifacts of the kind demonstrated by Goldiamond and Hawkins (1958); (*b*) when words were tachistoscopically presented in pairs, *S*s would look at one and then, if the exposure speed allowed, at the other, and this would result in lower thresholds for identification of either member than for discrimination; (*c*) the forced-response method would produce lower thresholds for word identification than would the free-response method; (*d*) reducing the size of the stimulus population by giving *S*s a list of the words to be shown would result in reduced identification thresholds; and (*e*) presenting a given word repeatedly at gradually increasing tachistoscopic durations would, by providing a series of pertinent fragmentary impressions, result in lower identification thresholds than would randomly presenting a group of words at each exposure speed (cf. Wyatt & Campbell, 1951).

METHOD

Materials.—Two sets of five-letter words were chosen from *The Teacher's Word Book* (Thorndike & Lorge, 1944). The words were randomly selected, with the stipulation that they be impressionistically neutral in tone, not closely similar in structure to previously selected words, and of greater than 400 or less than 10 per million frequency of occurrence. Within both of the six-word sets, each low frequency word was paired with a high frequency word, making three pairs of words per set (Set A: ADEPT, 8 occurrences per million, and MUSIC, 640 occurrences; GIRTH, 9, and SHAPE, 405; and INSET, 8, and DREAM, 745. Set B: CRIMP, 8 occurrences per million, and HEAVY, 823 occurrences; JAUNT, 8, and SPOKE, 946; and PLUSH, 4, and BEGAN, 1,928).

The words so chosen were printed by electric typewriter in pica capital letters on 8½ × 11 in. unwatermarked white bond paper. For single presentation, each word was centered on a separate sheet. For pairwise presentation, each pair was typed in all four possible combinations (e.g., ADEPT-MUSIC, MUSIC-ADEPT, ADEPT-ADEPT, MUSIC-MUSIC), one combination to a sheet. Each combination was centered with ¾ in. between the end of the first word and the beginning of the second (visual angle for *S* seated at apparatus = 0°53′40″).

The words were presented to *S*s in a mirror tachistoscope built by Merle Ridgley of the University of Illinois. Illumination of the pre-exposure field was .62 ft-c and of the exposure field 6.25 ft-c.

Subjects.—The *S*s were 104 male undergraduates at the University of Illinois who participated in the study as part of their course in introductory psychology. Each *S* was assigned unsystematically to one of the experimental conditions by a departmental secretary unfamiliar with the study.

Design.—The *S*s were seen individually by *E* under one of four experimental conditions (groups). In Group I ($N = 24$), *S*s were shown all 12 words, one at a time tachistoscopically, with the order of Word Sets A and B counterbalanced so that each set was presented first to half of the *S*s. Within each set, individual word order was also counterbalanced. Each word was exposed repeatedly at gradually longer durations (repeated presentation), and *S*'s task was to identify it. Half of the *S*s in Group I were required to make a response for each word presentation (forced response), and half were told to respond only if they saw a word (free response). No word lists were available to *S*s of Group I.

Group II ($N = 20$) differed from Group I only in the method of presenting the words. At a given exposure speed, all the words of the particular set were shown one after another in random order; at the next slower speed, all the words of the set were shown in a different random order, etc. (random presentation).

Group III ($N = 20$) differed from Group II only in that an alphabetical list of the words in each set was given to *S* before the word set in question was tachistoscopically presented.

Group IV *S*s ($N = 40$) were shown the words tachistoscopically in pairs, with four combinations for each pair of words. Half of the *S*s were shown only the 12 combinations of Set A words; half were shown only the 12 combinations of Set B. Presentation was random; at a given exposure speed, each of the 12 word combinations was exposed a total of two times. The *S*s were required for one series of presentations to identify both words in every two-word combination. For another separate series of presentations at the same exposure speeds, *S*s were required to tell whether the two words were the same or different. The order of these two tasks was counterbalanced across *S*s. All of Group IV was seen under the forced-response method, and each *S* was given an alphabetical list of the words of whichever set was being shown to him.

Procedure.—Each *S* was seated at the eyepiece of the tachistoscope with a plywood panel painted flat black separating him from *E*. The instructions below, modified as indicated according to *S*'s group and task, were read verbatim.

> What I am going to do is show you ([Groups I–III] some; [Group IV] pairs of) words on the screen in front of you. The words will appear ([Groups I–III] one at a time in the middle of the screen near; [Group IV] one to the left and one to the right of) that black dot. If you'll look now, I'll show you a sample so you can get an idea of what these will look like. [For Groups I–III, *E* then presented the sample word READY at an exposure speed of more than 500 msec. and asked *S* to identify it. For Group IV, *E* similarly presented the sample pair READY-TRACE.]
>
> OK. What I'm trying to find out here is how well you can see the words, and to do that I'll start at a very fast speed, show a bunch of these, and then get gradually slower and slower. ([For the repeated presentation of Group I, *S* was told instead] I'll start at a very fast speed, show you a word, and then keep repeating it slower and slower until you get it right several times in a row. Then we'll start over with a new word. [At this point, *E* said to *S*s of Groups III and IV] Here is a list of the words I'll be using. No other words will ever appear.)
>
> What I'd like you to do is to tell me each time ([Groups I–III] what the word was: [Group IV identification task] what the words were—first the one on your left and then the one on your right; it may help you to know that sometimes the two words will be the same—for example, "ready" on the left and "ready" again on the right—and sometimes they will be different; [Group IV discrimination task] whether the two words were the same—for example, "ready" on the left and "ready" again on the right—or whether they were different).
>
> I've tried to arrange this so that you probably won't see anything at all for the first few speeds. What I'd like you to do if you're not sure is to just take a guess; any word ([Groups I and II] at all; [Groups III and IV] from the list I gave you) will do. ([The free response *S*s of Groups I–III were told instead] What I'd like you to do if you're not sure is to just say "nothing.")
>
> Each time I'm ready to flash a ([Group IV] pair of) word(s) back here, I'll say "ready." You don't need to answer in any way; that will just be your signal that I'm about to show the word(s).

The *S*s were informed prior to each change in exposure speeds (Groups I–IV), word sets (Groups I–III), individual words (Group I), or tasks (Group IV identification to or from Group IV discrimination).

For Groups I–III, successive exposure speeds were 4, 9, 14, 21, 25, 39, 50, 60, 68, 76, 86, 96, 156, 240, 310, and 367 msec. Presentations of an individual word were terminated for Group I when three consecutive correct identifications of it had been made. Presentations of a particular word set to Groups

II and III were terminated when all six words were correctly identified at an exposure duration slower than 25 msec.

For Group IV, successive exposure speeds were 2, 4, 7, 11, 13, and 16 msec. Every *S* in this group was shown the words at all six speeds, regardless of the accuracy of his performance.

The *S*s in Groups I–III were run by a female research assistant. Half of the *S*s in Group IV were run by a male graduate student and half by the author.

RESULTS

High vs. low word frequency.—For *S*s in Groups I–III, the threshold for a word was defined as the exposure speed at which the first correct identification occurred. Data from the two no-list groups (I and II) were placed in mixed design analyses of variance (Lindquist, 1953) with word frequency, forced vs. free response, and groups (i.e., random vs. repeated presentation) as classifications. Separate analyses were done for each word set. The effect of word frequency on identification threshold was significant in both instances, Set A: $F\ (1,\ 40) = 34.96,\ p < .001$; Set B: $F\ (1,\ 40) = 11.92,\ p < .005$. The same data from the two random presentation groups (II and III) were placed in another pair of analyses of variances with word frequency, forced vs. free response, and groups (i.e., list vs. no list) as classifications. The main effect due to word frequency was again significant, Set A: $F\ (1,\ 36) = 27.92$; Set B: $F\ (1,\ 36) = 20.55$; p's $< .001$, but inspection of the data with respect to a Word Frequency × Group interaction, Set A: $F\ (1,\ 36) = 37.76$; Set B: $F\ (1,\ 36) = 22.40$; p's $< .001$, showed that the word frequency effect was significant only in Group II. Providing *S*s with a list of the words to be shown (Group III) was sufficient to eliminate the differential effects of high and low word frequency on the threshold for first identification.

When the above analyses for Groups I–III were repeated, using as data the exposure speed at which the initial *letter* of a word was first identified, no word frequency effects were shown for any group for either set of words (p's $> .05$). These results would seem to fit well with the notion that the relationship between word frequency and word thresholds is a function of *S*'s tendency to "guess" frequent rather than infrequent words to fill in fragmentary precepts.

For the identification task of Group IV, the total numbers of correct word identifications at each exposure speed were placed in mixed design analyses of variance with word frequency, exposure speeds, and *E*s as classifications. Separate analyses for Word Set A and B *S*s failed to show any effect of word frequency upon accuracy of identification (p's$>.05$). Three-way analyses of variance (*S*s × Exposure Speeds × Word Frequency), using as data the number of responses, regardless of correctness, further showed no high vs. low frequency response bias for either Set A or Set B *S*s (p's $> .05$). These findings for Group IV, which, like Group III, was provided with a list of the words, tend to support further the supposition that the frequency-threshold relationship depends upon response bias, which the list eliminated.

For the discrimination task of Group IV, only the word combinations where both words were identical could be used, since all other combinations contained both a high and a low frequency word. However, due to the fact that same discriminations were much more difficult than were different discriminations, at none of the exposure speeds used was the number of correct same responses significantly greater than chance expectancy for either word set, even though discrimination thresholds were shown. Nor were there any differences in the accuracy of same discriminations from any one exposure speed to another. Therefore, data from the identification task of Group IV were converted into same-different responses (by considering a response of, e.g., ADEPT-MUSIC as a "different" response, ADEPT-ADEPT as a "same" response, etc.). The number of such correct same responses at each exposure speed were placed in analyses of variance with word frequency, exposure speeds, and *E*s as classifications. Separate analyses of Word Set A and B *S*s showed no effect of word frequency upon discrimination of sameness (p's $> .05$), despite the fact that overall same discrimination thresholds were shown. It is, of course, still possible that if actual same discrimination thresholds had been reached, differences between high and low frequency words would have been discernible.

In summary, then, the analyses of Groups I–IV failed to reveal any differences in threshold between high and low frequency words which were not artifacts of differential response bias.

Discrimination vs. identification tasks.—For the purposes of this analysis, the number of correct word identifications at each of the six exposure speeds was determined for every Group IV *S*. The identification threshold for a given *S* was arbitrarily defined as the speed at which his accuracy first exceeded chance expectancy without decreasing to a chance or lower level at subsequent longer durations. A discrimination threshold for each Group IV *S* was similarly determined from the accuracy of his same-different responses. Separate t tests were done for the Word Set A and B *S*s to test the hypothesis that identification thresholds are lower than discrimination thresholds. The t values of 4.77 for Set A *S*s and 3.91 for Set B *S*s were beyond the .0005 level of significance for the one-tailed test with $df = 19$.

The Group IV *S*s repeatedly commented that at certain speeds they could see only one of the two words. This observation was tested more objectively by tallying for each Group IV *S* the number of correct and incorrect identifications of the left and right hand words. For each *S* a χ^2 was computed to see if correct identification of one word in a pair was associated with incorrect identification of the other word. These χ^2 values, summed across *S*s, were significant beyond the .001 level for both word sets, $\chi^2\ (20) = 106.90$ for Set A *S*s and 86.34 for Set B *S*s. Thus, correct identification of one of the words in a pair was negatively correlated with correct identification of the other, presumably because the exposure speeds were generally not slow enough for *S* to see both words.

Forced- vs. free-response methods.—For Groups I–III, all of the analyses described above to test for frequency effects on first identification thresholds included forced vs. free response as a classification. For identification of the entire word, no forced- vs. free-response effects were found in any of Groups I–III for either set of words ($p > .05$). For identification of the initial letter of the word, forced- vs. free-response effects appeared for Groups II and III for Word Set B only, $F\ (1, 36) = 4.15$, $p < .05$, with forced-response *S*s identifying the initial letter at faster exposure speeds.

When analyses were done of the number of correct identifications across exposure speeds (i.e., threshold curves), the expected effects were more clearly present. Mixed design analyses of variance were done for each group separately, using *S*s, exposure speeds, and forced vs. free response as classifications. For Group I, there were no differences between the forced- and free-response methods for either the Word Set A or B *S*s (p's $> .05$). For Group II, no main effects due to the forced- vs. free-response methods appeared for either Set A or Set B *S*s (p's $> .05$), but a Forced vs. Free Response × Exposure Speed interaction for the Word Set B *S*s resulted from the fact that the free responses were more accurate at the slowest of the seven exposure speeds on which there were data for all *S*s, $F\,(6,\ 108) = 2.74$, $p < .025$. For Group III, significant forced- vs. free-response main effects were shown for both Set A and Set B *S*s, Set A: $F\,(1,\ 18) = 6.00$, $p < .025$; Set B: $F\,(1, 18) = 5.92$, $p < .05$, with forced responses being more accurate.

Generally, then, analyses of first identification speed did not show consistent differences between forced- and free-response *S*s, while analyses of accuracy of identification across exposure speeds showed consistent forced- vs. free-response differences for Group III only.

Population size (list vs. no list).—The threshold differences between the list and no list (random presentation) Groups II and III were consistently significant, Set A: $F\,(1, 36) = 19.54$; Set B: $F\,(1, 36) = 22.69$; p's $< .001$, using the first identification speed data. However, different chance expectancies for the two groups made this finding questionable. These differing chance levels also precluded analyses of variance of accuracy across exposure speeds. Therefore, the population size hypothesis was tested by determining for each *S* in the two groups the accuracy of identification at each exposure speed. The threshold for a given *S* was arbitrarily defined as the speed at which his accuracy first exceeded chance expectancy without decreasing to a chance or lower level at subsequent slower speeds. (For Group II, no list, chance expectancy was assumed to be zero correct identifications out of the six word presentations at each exposure speed. It should be noted that incorrectness of this assumption would work against the hypothesis to be tested: that no-list thresholds are higher than list thresholds.) These threshold data were placed in two-way analyses of variance (Forced vs. Free Response × Groups), one analysis for each word set. The effect of population size was shown in a significant difference between Group II, no list, and Group III, list, Set A: $F\,(1,\ 36) = 6.11$, $p < .025$; Set B: $F\,(1,\ 36) = 17.91$, $p < .001$. Inspection of the data with respect to a significant two-way interaction, Set A: $F\,(1,\ 36) = 5.98$, $p < .025$; Set B: $F\,(1, 36) = 4.14$, $p < .05$, showed that curtailing the stimulus population by giving *S*s a word list reduced overall identification thresholds only for forced-response, and not for free-response *S*s.

Random vs. repeated presentation.—The first identification speed analyses for the initial letter and for the whole word showed no differences due to random vs. repeated presentation (Group I vs. Group II) for either word set (p's $> .05$). The same absence of differences was found for both word sets when identification accuracy across exposure speeds was placed in mixed design analyses of variance with *S*s, exposure speeds, and groups (i.e., random vs. repeated presentation) as classifications. It was concluded that manner of presentation made no difference. This result is a failure to replicate the work of Wyatt and Campbell (1951), who found that the ascending method of limits (i.e., repeated presentation) produced higher thresholds than random presentation because *S*s clung to erroneous hypotheses across trials.

Experimenter effects.—As noted above, various analyses included *E*s as a classification. In no case were there differential effects on the findings.

Order effects.—A number of analyses of variance were performed to test for effects of word set order (Groups I–III), word order within sets (Group I), and task order (Group IV), upon accuracy of identification. No such effects were observed.

DISCUSSION

The present results lend support to previous findings of Goldiamond and Hawkins (1958) regarding the artifactual nature of the frequency-threshold relationship. Although there was no difference between high and low frequency words in the exposure speed at which the initial letter was correctly identified, differences appeared in the speed of first identification for the entire word. This presumably occurred because *S*s guess frequent rather than infrequent words to fill in fragmentary tachistoscopic percepts and so tend to be correct earlier in naming frequent words. When *S*s were given a list of the six words to be tachistoscopically presented, the difference in identification thresholds between high and low frequency words disappeared, since fragmentary percepts (e.g., the first letter) were then sufficient to enable *S*s to identify a word.

Threshold differences between high and low frequency words in the absence of a word list were shown even for the free-response method *S*s, who were told not to respond if they were uncertain.

It is clear that *S*s are not able to discriminate between two words in a spatial pair at a faster exposure speed than is required for identification. Instead of approaching the word pair as a unit, they apparently look first at one of the words and then, if the exposure time allows, at the other. As a result, they are able at some speeds to identify one word without getting enough information about the other to make a nonchance same-different response. This finding suggests that when groups of stimuli are simultaneously presented (e.g., Blum, 1954), tasks involving discrimination among the stimuli probably require longer tachistoscopic exposure durations than do identification tasks. It is also suggests that asking *S*s which stimulus "stands out most" (Blum, 1954) amounts to asking which stimulus is seen. Unless there are threshold differences between the stimuli, position habits would probably determine *S*s' clearness responses, and as a result perceptual vigilance could not be shown. Of course, when the stimuli are presented successively (e.g., Klein, 1959; Taylor, 1953), it would still be expected that discriminations among stimuli could be made at speeds below the identification threshold.

Just as supplying *S*s with a word list produced marked changes in the relationship between high and low frequency word thresholds, it also affected the relationship between the forced- and free-response methods. When no list of words was available, forcing *S*s to respond had no consistent effect upon accuracy of identification. However, when *S*s had a word list, the forced-response method resulted in increased accuracy. It appears that the forced-response method is superior only when it can lead *S*s to use additional available information in their judgments.

The above finding fits well with Eriksen's (1958) discussion of the "subjective confidence interval," for it implies that *S*s set some arbitrary personal standard for when they will report a word. Reducing the size of the stimulus population by supplying a word list provides additional pertinent information, but only *S*s who are forced to do so make use of it.

This finding concerning population size and thresholds also has relevance for studies which show that forewarnings of the taboo nature of the stimuli produce lowered thresholds for "dirty" words (e.g., Freeman, 1954). In many instances, the results of these preparatory set experiments have been interpreted to mean that perceptual defense arises from response suppression—that is, *S*s "see" the dirty words but hesitate to report them. It is obvious, however, that preparatory set simultaneously reduces the size of the stimulus population. Thus, *S*s who have been prepared for taboo words have more information relevant to, for example, the percept SH-T than do unprepared *S*s and are therefore more likely to identify it as a taboo word. Likewise, it would seem unnecessary to postulate perceptual vigilance to threat conditions when *S*s prepared for obscenity show lower thresholds for taboo words than for interpolated innocuous words.

REFERENCES

BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *U. Mich. Engng. Res. Inst. Bull.*, 1953, No. 36.

BLUM, G. S. An experimental reunion of psychoanalytic theory with perceptual defense and vigilance. *J. abnorm. soc. Psychol.*, 1954, **49**, 94–98.

BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69–77.

COWEN, E. L., & BEIER, E. G. Threat-expectancy, word frequencies, and perceptual prerecognition hypotheses. *J. abnorm. soc. Psychol.*, 1954, **49**, 178–182.

ERIKSEN, C. W. Unconscious processes. In M. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958. Pp. 169–227.

ERIKSEN, C. W., & BROWNE, C. T. An experimental and theoretical analysis of perceptual defense. *J. abnorm. soc. Psychol.*, 1956, **52**, 224–230.

FREEMAN, J. T. Set or perceptual defense? *J. exp. Psychol.*, 1954, **48**, 283–288.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.

KLEIN, G. S. On subliminal activation. *J. nerv. ment. Dis.*, 1959, **128**, 293–301.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.

POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.

PUSTELL, T. E. The experimental induction of perceptual vigilance and defense. *J. Pers.*, 1957, **25**, 425–438.

TAYLOR, F. W. R. The discrimination of subliminal visual stimuli. *Canad. J. Psychol.*, 1953, **7**, 12–20.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

WIENER, M. Word frequency or motivation in perceptual defense? *J. abnorm. soc. Psychol.*, 1955, **51**, 214–217.

WYATT, D. F., & CAMPBELL, D. T. On the liability of stereotype or hypothesis. *J. abnorm. soc. Psychol.*, 1951, **46**, 496–500.

(Received October 29, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 371–376

DISCRIMINATION OF REWARDS AS A FUNCTION OF CONTRAST IN REWARD STIMULI[1]

MARK A. BERKLEY[2]

Johns Hopkins University

104 Ss were given a .006-cc drop of a 10 gm/l saccharine solution every 11th lick from a drinking tube. On the intervening 10 licks, Ss received nothing, or they received fluids that varied on 2 dimensions: saccharine concentration and volume. The results show that decreasing the concentration of the fluid delivered on the 10 intervening responses significantly increased resistance to extinction. The volume variable had no significant effect on either licking rate during training or on resistance to extinction. The relevance of these findings to the effect of magnitude of reward on resistance to extinction after training with partial reinforcement was discussed.

Recently, it has been shown (Hulse, 1962a, 1962b) that partial reinforcement (PR) provides *S*s with reward discrimination training by enabling them to contrast the stimuli associated with reward on reinforced responses with the stimuli of nonreward on nonreinforced responses. With continuous reinforcement (CR) this cannot occur since the stimuli correlated with the reinforcement are the same on every response. Partial reinforcement thus allows *S*s to respond to qualitative aspects of the reward while CR does not.

With the discrimination process in PR, the contrast between the stimuli associated with a given magnitude of reward and nonreward should be greater for *S*s trained under PR than to *S*s trained under CR. If this contrast of stimuli is correlated with the reinforcement value of the reward, the reinforcement value of a reward given on a PR schedule may be greater than it is when the same magnitude of reward is given on a CR schedule. On the assumption that the reinforcement value of a reward can be measured by resistance to extinction, the discrimination hypothesis predicts that the greater the magnitude of reward employed during PR training, the greater should be the resistance to extinction. This effect should not be seen with CR training. A number of investigators have, in fact, reported such results (Hulse, 1958; Hulse & Bacon, 1962; Lewis & Duncan, 1961; Reynolds, 1950; Wagner, 1961).

If discrimination is taking place during PR, the contrast between stimuli associated with qualitative or quantitative aspects of reward (S_R) and the stimuli associated with nonreward (S_O) may be decreased by making S_O approximate S_R. Thus, for example, a fluid reward of constant volume (S_R) could be presented every tenth response. On the intervening nine responses, another stimulus (S_O), a smaller volume of fluid than that presented every tenth response, could be presented. Here, the S_R-S_O contrast should be reduced as compared to a situation like true PR, for example, where S_O is nothing at all and where the contrast between S_R and S_O is at a maximum. In general, the limiting cases are CR, where S_O is the same as S_R, and PR, where S_O is nonreward and S_R is any magnitude of reward.

---

[1] This paper is based upon a dissertation submitted in partial fulfillment of the requirements for the PhD degree. The author is indebted to Stewart H. Hulse, dissertation advisor, for his most helpful counsel. This research was supported in part by a National Science Foundation Grant, G18125, to Stewart H. Hulse.

[2] Now a National Institute of Mental Health Postdoctoral Fellow at Brown University.

In this study, behavior was examined when the contrast between S_R and S_O was systematically varied. The instrumentally-conditioned licking response was used. All *S*s received an S_R of a constant volume and concentration of saccharine every eleventh response. On the intervening 10 responses, *S*s were given other stimuli (S_O) with which S_R could be compared. The S_O stimuli were fluids which varied along two stimulus dimensions, saccharine concentration and volume.

METHOD

Subjects

The *S*s were 112 experimentally naive male albino rats of the Sprague-Dawley strain. They ranged in age from 80 to 120 days at the start of taming.

Apparatus

The apparatus consisted of a sound-resistant refrigerator box which had a 3-w. lamp attached to the inside of its lid. A wooden test box, 4.5 × 10 × 6 in., was placed inside this box. The test box was painted flat black and had a grid floor made of $\frac{1}{8}$-in. brass rods spaced $\frac{1}{2}$ in. apart. The hinged lid and one end of the test box were made of Plexiglas. The Plexiglas end had a 1 × $\frac{3}{8}$ in. vertical slot located in its center. A drinking tube was mounted behind this slot.

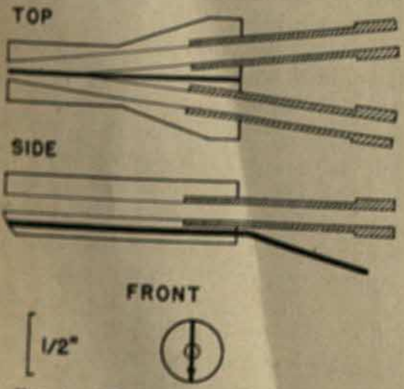


FIG. 1. Three views of the two-channel drinking tube. (The hatched areas represent stainless steel tubes which were connected to the fluid pumps.)

Three drinking tubes were used in this study: a $\frac{1}{8}$-in. brass training tube, a $\frac{1}{2}$-in. diameter Plexiglas single-channel tube (Hulse, 1960), and a two-channel tube. Both single-channel tubes had 2-mm. diameter fluid channels. The two-channel tube is shown in Fig. 1. It was made of $\frac{1}{2}$-in. round Plexiglas and had two channels through which fluids could be delivered. The two channels were separated by a vertical sheet of Plexiglas .010 in. thick. The semicircular opening of each channel had an area of 3.141 mm^2. A $\frac{1}{16}$-in. brass wire was located $\frac{1}{16}$ in. below the two-channel aperture to provide an electrical contact.

Each channel of the two-channel drinking tube was connected by pressure tubing to a fluid pumping system described by Hulse (1960). Each pumping system included a fluid pump and a solenoid-operated lever. An electronic relay which operated each time *S*'s tongue touched the drinking tube contact operated the pumping system through a programing circuit. This system allowed the delivery of fluid drops of different sizes and concentrations to the tip of the drinking tube.

Procedure

Experimental design.—All *S*s received a .006-cc drop of a 10 gm/l saccharine solution (S_R) on every eleventh response, but *S*s received different fluids (S_O) on the intervening 10 responses as determined by the following design. The design was a factorial which combined four saccharine concentrations, .00 gm. (water), .25 gm., 2.50 gm., and 10 gm. of saccharine per liter of tap water, with three volumes of fluid, .001 cc, .003 cc, and .006 cc. The design required 12 experimental groups ($N = 8$ in each group). Each group was designated by a letter corresponding to its drop size (S, M, or L) and a number designating its saccharine concentration (.00, .25, 2.50, or 10). For example, Group M-.25, like all groups, received a .006-cc 10 gm/l drop (S_R) every eleventh response. On the intervening 10 responses, it received a .003-cc drop of the .25 gm/l saccharine solution (S_O). It will be noted that Group L-10 was on a traditional continuous reinforcement schedule

(CR) where $S_R = S_O$. Both volume and concentration were constant from response to response. Similarly, all L groups were on a CRF schedule with respect to volume and all 10 groups were on a CRF schedule with respect to concentration.

In addition to the groups called for by the design, two control groups were run. One (Group W-C) was continuously reinforced with .006 cc/lick water and the other (Group P-C) was partially reinforced with .006 cc of a 10-gm/l saccharine solution every eleventh response, but nothing on the intervening 10 responses.

Taming and deprivation.—The *Ss* were housed in large colony cages for 21 days after arrival from the supplier. Ad libitum food and water (in troughs) were available. At the end of this period, *Ss* were transferred to individual cages and placed on a deprivation schedule. Each *S* received a daily ration of 10 gm. ground Purina chow mixed with 15 cc of tap water.

Taming consisted of placing groups of 10–15 *Ss* in large boxes for 10 min. each day. They were allowed to explore and were freely handled by *E*. The daily food and water ration was given upon *Ss'* return to the home cage. This procedure continued for 27 days.

Training.—Following taming, *Ss* were given 30 training sessions, 1 session per day. A day's session consisted of 400 licks or 3 min. in the test box, whichever occurred first. After a day's session, *Ss* returned to their home cages.

On Day 1, the brass tube protruded through the slot in the Plexiglas wall. It was withdrawn a small amount on subsequent days so that on Day 4 its tip was located slightly behind the slot. On Days 5–8, the brass tube was replaced by the Plexiglas single-channel tube. For Days 9–30, the two-channel tube was used.

All *Ss* were continuously reinforced with .006 cc per lick of the 10 gm/l saccharine solution (S_R) from Day 1 to Day 12. From Days 13–17, the S_O-S_R schedules were gradually introduced. On Day 13, the *Ss* received S_R on every other response. On the intervening responses, each of the 12 groups received an S_O as specified by the design. On Day 14, the S_O-S_R ratio was increased to 3:1, on Day 15 to 5:1, on Day 16 to 8:1, and on Day 17 to 11:1, the final schedule for the remainder of the training sessions. The order in which the groups were run each day was randomized; so was the order of running the *Ss* within each group.

After a day's session, *Ss* returned to their home cages where they received their ration of wet mash. The ration was given each day so that *Ss* would be 22–24 hr. food and water deprived at the start of the following day's session. During the course of the experiment, the water ration was adjusted to account for the fluid received in the test apparatus. The total fluid intake always equaled 15 cc per *S* per day.

TABLE 1

MEANS AND *SDs* FOR LICKING RATES DURING TRAINING

Group	*M*	*SD*
P-C	4.087	0.878
.00	4.704	0.649
.25	4.488	0.535
2.50	4.751	0.481
10	4.863	0.310

Note.—*N* = 8 in Group P-C; *N* = 24 in the other groups.

Extinction.—Four days of extinction were run. The procedure was the same as during training except that no fluid was placed in the pumps. The *Ss* thus licked on the dry two-channel tube. For each session, *S* was allowed 2 min. in the test box.

Measures of performance.—During training, the time each *S* required to emit 400 licks was recorded to the nearest second. An average licking rate, in licks/sec, was then computed for each of the last four training sessions for each *S*.

During extinction, the number of responses each *S* made during each extinction session was recorded.

RESULTS

Table 1 shows the mean licking rates and *SDs* for the S_O concentration groups. Among the experimental groups, the fastest licking rate during training was correlated with the 10-gm. S_O. The slowest licking rate was correlated with the .25-gm. S_O, with the other two S_O concentrations falling in between. The group having the slowest average licking rate was the P-C control group.[3]

An analysis of variance of the

---

[3] The data of the W-C control group were not analyzed because these *Ss* never reached the 400-lick criterion during training. This may have been due to insufficient deprivation.

TABLE 2

Analysis of Variance of Licking Rates during Training

Source	*df*	*MS*	*F*
*S*s	95		
Concentration (C)	3	2.374	3.311*
Volume (V)	2	1.884	2.628
C × V	6	0.243	0.339
Error (within)	72	0.717	
Days (D)	3	0.386	3.328*
D × *S*s	285		
C × D	9	0.099	0.853
V × D	6	0.124	1.069
C × V × D	18	0.098	0.845
Error (within × D)	252	0.116	

* $p < .05$.

licking data (Table 2) showed that concentration of So produced significant differences in licking rates among experimental groups ($p < .05$). However, the So variable of volume per lick did not produce any significant differences. Although the overall effect of concentration was significant, a multiple-range test (Duncan, 1955) failed to show any significant differences between pairs of individual group means.

The last 4 days of training were also treated as a separate variable in the analysis of variance. This variable was found to be significant ($p < .05$) which indicates that *S*s had not reached an asymptotic performance level at the end of training. The licking rates were apparently increasing in equal amounts across all groups, since none of the interactions involving days was significant.

The extinction results are shown in Fig. 2. The figure shows that, of the experimental groups, the one most resistant to extinction was Group .00. This result reflects a significant concentration effect ($p < .001$) shown in Table 3. A multiple-range test showed that Group .00 was significantly different from Group .25 and Group 10 ($p < .05$) but not from Group 2.50. Group P-C showed the greatest overall resistance to extinction. The difference in the total number of responses emitted by *S*s in Group P-C and Group .00 was significant, $t(30) = 2.01$, $p < .05$, one-tailed.

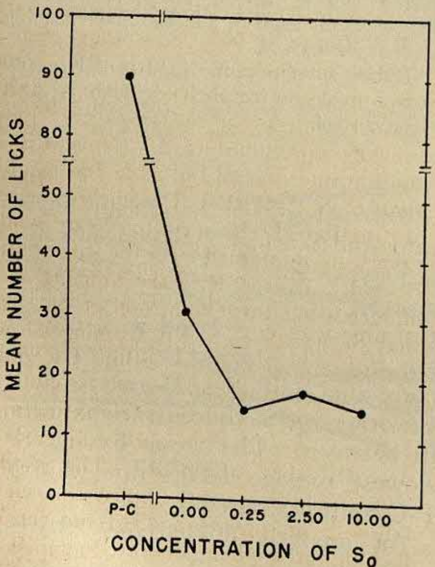


Fig. 2. Mean number of licks emitted during extinction for groups with different concentrations of So, and for the P-C control group.

TABLE 3

Analysis of Variance of Number of Licks during Extinction

Source	*df*	*MS*	*F*
*S*s	95		
Concentration (C)	3	5778.51	6.92***
Volume (V)	2	1771.27	2.12
C × V	6	1564.32	1.87
Error (within)	72	834.91	
Days (D)	3	65275.06	162.46***
D × *S*s	285		
C × D	9	1425.31	3.55***
V × D	6	352.12	0.876
C × V × D	18	329.42	0.820
Error (within × D)	252	401.78	

*** $p < .001$.

A highly significant days effect was

also found during extinction ($p < .001$) and indicates that on Day 1 of extinction, all groups produced a greater number of responses than on subsequent days. Inspection of group means suggested that the significant interaction between extinction sessions (Days) and concentration (Table 3) appeared to be due to the difference in rate of extinction of *S*s in Group .00 as compared to the other experimental groups. Although *S*s in Group .00 extinguished at a faster rate, they emitted the greatest number of responses on all extinction days, and never reached the level of the other groups. None of the other main effects or interactions was found to be significant.

Discussion

Resistance to extinction increases as S_R becomes more discriminable from S_O. For the present data, this effect is clearly shown for the .00 S_O concentration and for the P-C control groups. For these conditions, the S_R-S_O contrast was greatest.

With PR, the contrast between S_R and S_O can be increased in two ways: increasing the magnitude of S_R and holding S_O constant, or by holding S_R constant and decreasing S_O. The former procedure was used in a study by Hulse and Bacon (1962) where S_O was nonreward, and the concentration of a sacchrine reward (S_R) was increased. The latter procedure was used in this study where S_R was held constant and S_O systematically varied. In both experiments resistance to extinction varied with S_R-S_O contrast. Both studies also showed that the magnitude of reward had no effect on resistance to extinction after CR training.

A number of experiments show that varying the magnitude of reward from response to response during training increases resistance to extinction (Logan, Beier, & Kincaid, 1956; Wike, Kintsch, & Gutekunst, 1959; Wunderlich, 1961; Yamaguchi, 1961). The approach in these experiments is similar to the present approach in that *S*s were able to compare one magnitude of reward with another. In these studies, however, the stimuli to which *S*s were responding are not clear. The *S*s could have been responding to quality of reward, or to cues correlated with the different number of consummatory responses required for each magnitude of reward since these earlier experiments used different numbers of food pellets or weights of wet mash as rewards. In the present study, variation in the stimuli associated with consummatory responding was minimized because the consummatory and operant responses were the same. Consequently, the stimuli most clearly free to vary were those associated with the quality or quantity of the reward. The data indicate that when consummatory response stimuli are essentially constant, as is the case in the present study, quality as compared with quantity of the reward has a greater effect on the contrast of the reward stimuli.

An extension of the discrimination hypothesis may explain, at least in part, the greater resistance to extinction found in *S*s trained with PR as compared with CR. Logan (1960) has suggested a similar mechanism. With PR, some responses are reinforced with a finite reward while other responses are "reinforced" with nonreward. Thus, PR is the limiting case of reducing the magnitude of S_O. With CR, some responses are reinforced with a finite reward while others are reinforced with the same finite reward. Thus, CR is the limiting case of increasing the magnitude of S_O such that S_O equals S_R. PR training allows reward-stimuli contrast while CR does not. If contrast of reward stimuli is correlated with resistance to extinction, resistance to extinction should be greater after PR than after CR. The results of this study tend to support this view.

Finally, it appears that S_R is most easily discriminated when presented against an S_O of water or an S_O of nonreward. The other concentration used in this study had little effect on resistance to extinction. This may have occurred because the .25 S_O

and the 2.50 S_O were not discriminable from the 10 S_R, or because there was not a sufficient number of S_O responses between S_R responses.

REFERENCES

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

HULSE, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

HULSE, S. H. A precision liquid feeding system controlled by licking behavior. *J. exp. anal. Behav.*, 1960, **3**, 1–3.

HULSE, S. H. Discrimination of the reward in learning with partial and continuous reinforcement. *J. exp. Psychol.*, 1962, **64**, 227–233. (a)

HULSE, S. H. Partial reinforcement, continuous reinforcement, and reinforcement shift effects. *J. exp. Psychol.*, 1962, **64**, 451–459. (b)

HULSE, S. H., & BACON, W. E. Supplementary report: Partial reinforcement and amount of reinforcement as determinants of instrumental licking behavior. *J. exp. Psychol.*, 1962, **63**, 214–215.

LEWIS, D. J., & DUNCAN, C. P. Effects of variable magnitude of reward on a lever-pulling response. *J. exp. Psychol.*, 1961, **62**, 203–205.

LOGAN, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

LOGAN, F. A., BEIER, E. M., & KINCAID, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, **52**, 65–70.

REYNOLDS, B. Resistance to extinction as a function of the amount of reinforcement present during acquisition. *J. exp. Psychol.*, 1950, **40**, 46–52.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

WIKE, E. L., KINTSCH, W., & GUTEKUNST, R. The effects of variable drive, reward, and response upon instrumental performance. *J. comp. physiol. Psychol.*, 1959, **52**, 403–407.

WUNDERLICH, R. A. Strength of a generalized conditioned reinforcer as a function of variability of reward. *J. exp. Psychol.*, 1961, **62**, 409–415.

YAMAGUCHI, H. G. The effect of continuous, partial, and varied magnitudes of reinforcement on acquisition and extinction. *J. exp. Psychol.*, 1961, **61**, 319–324.

(Received October 30, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 377–382

EFFECT OF AMOUNT AND DISTRIBUTION OF INSPECTION TIME AND LENGTH OF DECAY INTERVAL ON KINESTHETIC AFTEREFFECT[1]

JEAN B. CARLSON[2]

Northwestern University

This was a study of the effect of massing or distributing increasing amounts of inspection time on the magnitude and rate of dissipation of kinesthetic aftereffect. Ss were given pretest (1-in. test block), then massed (continuous) or distributed inspection (1-min. inspection periods separated by 2 min. rest) with a 2-in. block, and finally a posttest immediately, 1, or 7 days, after inspection. Results were: magnitude of aftereffect increased as a function of amount of inspection time, and decreased over rest; magnitude of aftereffect did not differ as a function of massing or distributing inspection; and there was no interaction between decay rate and duration of inspection.

In Köhler and Wallach's (1944) study of figural aftereffects, and in review articles by Duncan (1956), McEwen (1958), and Sagara and Oyama (1957), it was suggested that magnitude of aftereffect may reach an asymptote at some point during single continuous inspection (I) periods. These authors suggested that additions in I time beyond this point would not create a measurably greater amount of immediate aftereffect (i.e., when presentations of the test—T—object immediately follows presentation of the I object), but might affect the rate of dissipation of aftereffect and thus create a more enduring effect. The first problem of the present study was to examine the possibility that long I times would affect the rate of dissipation of (kinesthetic) aftereffect more than the initial magnitude of effect.

The second problem of this study grew out of findings of Köhler and Dinnerstein (1947), Wertheimer and Leventhal (1958), and Charles and Duncan (1959) on kinesthetic aftereffects. All of these investigators reported that if *S*s were given brief exposures to an I object on each of a number of successive occasions, magnitude of initial aftereffect increased after each presentation of the I object. In the first two studies cited, the T object-I object-T object sequence was repeated at 24-hr. intervals (for a month and a half by Köhler and Dinnerstein, for 5 days by Wertheimer and Leventhal). In the Charles and Duncan study there were only two sequences which were separated by 2 min. of rest. These studies suggest that the size of initial aftereffect may be increased when I time is increased by giving several I periods (distributed inspection) rather than by increasing the length of a single continuous inspection period (massed inspection). In these three studies, however, it is not possible to determine whether the increase in aftereffect following an increase in the number of inspection periods was due to the distribution of I time or to the greater total amount of I time.

---

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at Northwestern University. The author wishes to acknowledge her indebtedness and express her appreciation to Carl P. Duncan for his guidance, advice, and criticism throughout the course of this study.

[2] Now at Kendall College, Evanston, Illinois.

Further, it is possible that repeated testing may have influenced the magnitude of aftereffect since a test with a T object was given after each I period in all cases. Therefore, it was the second purpose of the present study to determine the effect of distributing increases in total I time, without intervening T periods, on the size and persistence of kinesthetic aftereffect.

METHOD

Materials.—Experimental materials were taken from those constructed by Charles and Duncan (1959). They consisted of two wooden blocks and a wedge. The blocks were 1 ft. long and 2 in. high. One block was 1 in. wide and was used as the T block. The other block was 2 in. wide and was used as the I block.

Judgment of the width of the T block was made on a wedge of wood 40 in. long, 2 in. high, and increasing in width from 0 to 4 in. at the rate of 1 in. increase in width for each 10 in. of length. The vertical sides of the blocks and wedge were covered with smooth plastic coated paper. A tape measure, graduated in $\frac{1}{8}$-in. units, was glued to the top of the wedge. An iron rod mounted $\frac{3}{4}$ in. above the wedge supported a plastic H shaped slider which, when gripped by *S*, insured that the thumb and finger were in a single plane at right angles to the center line of the wedge. The slider also permitted accurate reading of the scale.

Blocks were held rigid at both ends in a block holder. The block holder and wedge scale were set on tables parallel to each other and about 20 in. apart. A tape measure fixed to the side of the table permitted accurate location of the scale on the table.

Design.—Groups of experimental *S*s were given the I block for 1, 3, or 5 min. of massed inspection, or 3 or 5 min. of distributed inspection (1-min. periods of inspection separated by 2 min. of rest). Each of the five major groups (three massed and two distributed) was divided into three subgroups. One subgroup was tested immediately after inspection, a second was tested 1 day after inspection, and a third was tested 7 days after inspection. Thus, there were 15 experimental subgroups. No *S* was tested more than once.

Two control groups were run which were given Practice, a Pretest, and Posttest (see Procedure) but no I period. One control group rested 13 min. between Pretest and Posttest (corresponding to the 5-min. distributed inspection-immediately tested group). This represents the longest interval between Pretest and Posttest for any *S* who completed the experiment within 1 day. The other control group had 7 days between Pretest and Posttest (the longest decay interval period intervening between Pretest and Posttest for any experimental group).

Subjects.—The *S*s were 225 students from introductory psychology courses. The *S*s were assigned to groups in turn until each group contained 12 *S*s after which assignment was made on the basis of Pretest scores to achieve rough matching of groups. The final *N* was 15 in each of 17 groups.

Procedure.—The *S*s were run individually. First *S* was told the experiment concerned the ability of humans to judge widths of blocks using the thumb and middle finger of each hand. Each *S* was shown by *E* how the nonpreferred hand should grip a block placed in the block holder while the preferred hand moved up and down the wedge scale until a width was found which was judged by *S* to be equal in width to the width of the block. After instruction, *S* was blindfolded and led to a standard position between two tables so that the wedge scale was presented to the side of his preferred hand and the block in the holder to his other hand. The *S* was assisted to grip the block with one hand and the wedge with the other. In finding the point of subjective equality, *S* was instructed to move back and forth over the range of apparent equality until the choice became narrowed to a particular width.

The wedge position on the table and the slider position on the wedge were systematically varied within 11 in. for the wedge and 10 in. for the slider so that at the start of each judgment, the wedge was at the back or front of the table and the slider back or forward of the true equality point on the wedge.

An *S* was first given the 1-in. T block and eight judgments were made of this block. The first four judgments were called Practice trials during which *S* was encouraged to make judgments swiftly. The next four trials were called the Pretest trials. The mean of the Pretest trials was used as a PSE, before inspection, for each *S*. As soon as *S* made his setting of the slider on the wedge, a trial was completed. Between trials, *S* let go of both block and wedge and stood with hands at his sides. Time between trials was only that necessary for *E* to record the setting and move the wedge and slider for the next trial.

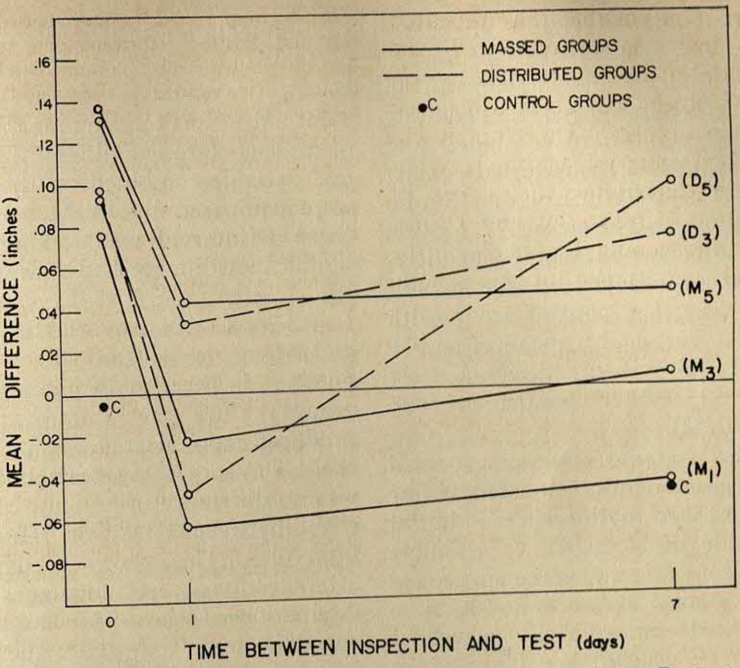


FIG. 1. Mean difference between mean of four Pretest trials and first Posttest trial for experimental groups given 1, 3, or 5 min. inspection under distributed (D) or massed (M) conditions, and for control (C) groups.

After the Pretest trials, all *S*s except the Control groups were given inspection with the 2-in. I block. Each *S* was asked to press fairly firmly against the sides of the block with the nonpreferred hand and then move the thumb and middle finger back and forth along its sides at a rate of approximately 100 moves per minute. The *S* was to continue until told to stop. In the groups given distributed inspection, 1 min. of inspection was followed by 2 min. of rest with the cycle repeated until the total I time of 3 or 5 min. was completed. During rest, *S*s sat down for 2 min. until the next I period.

After completion of inspection, *S*s were either given additional trials, called Posttest trials, immediately, or were dismissed to return after 1 or 7 days. During the Posttest trials, the 1-in. T block was again presented to the nonpreferred hand and *S* was asked to make four judgments of the size of the block.

RESULTS

Pretest.—The group means of the Pretest PSEs for the 17 groups differed by not more than .03 in., and the variances by not more than .012 in. A comparison of the highest and lowest mean gave a *t* of 1.00.

The grand mean of the Pretest PSEs for all 225 *S*s was 1.08 in.; thus the constant error of .08 in. (apparent overestimation of the T block) was in the opposite direction from that expected from aftereffect, which should show apparent shrinkage or underestimation of the size of the T block.

Aftereffect.—The measure of aftereffect for each *S* was derived by subtraction of his judgment on the first Posttest trial from his Pretest PSE. Subtracted this way, the presence of aftereffect is suggested by positive values, since by the nature of the conditions, the T block should shrink. The means of these difference scores for experimental and control groups are shown graphically in Fig. 1.

To determine if significant aftereffect occurred as the result of inspection of the I block, scores of all experimental *S*s tested immediately after inspection were combined and the mean (.11 in.) was tested against the mean (−.001 in.) of the 13-min. control group. The t was 2.77, $p < .01$; thus it appears that significant aftereffect was produced in the experimental *S*s. The control group with the 7-day rest had a mean of −.049 in. and did not differ significantly from the other control group ($\sigma_{\text{diff.}} = .52$).

Massed I groups.—It was expected that large amounts of massed I time would not lead to differences in initial magnitude of aftereffect. A simple analysis of variance of the subgroups given 1, 3, or 5 min. of massed I time and tested immediately suggested that there was little difference in initial amount of aftereffect in groups given more than 1 min. of I time; numerical differences appeared (Fig. 1) but the differences were not significant, $F(2, 42) = .82$.

If large amounts of massed I time affect rate of dissipation of aftereffect there should be an interaction between amount of I time and time of testing. This was not found; the interaction term was not significant in a two-way analysis of variance of the variables, amount of massed I time (1, 3, or 5 min.) and duration of the T interval (0, 1, or 7 days between inspection and test), $F(4, 126) = .12$.

In the two-way analysis both main variables were significant; magnitude of aftereffect increased as massed I time increased, $F(2, 126) = 3.99$, $p < .05$, and decreased over rest, $F(2, 126) = 8.26$, $p < .01$. The fact that the F for amount of I time was significant in this analysis, where it had not been in the analysis of only the immediately-tested subgroups, is a result of the larger N available for the two-way analysis.

Grand means for amount of I time of testing (groups differentiated on one variable combined on the other) were compared by t test. For the three T interval means, there were significant differences between the Immediate and 1-Day groups ($t = 3.74, p < .01$), and the Immediate and 7-Day groups ($t = 3.30, p < .01$), but not between 1-Day and 7-Day groups; thus the Immediate test produced more aftereffect than either the 1-Day or 7-Day tests but there was no difference in amount of aftereffect between 1-Day or 7-Day test intervals. This shows a decay of aftereffect over rest and suggests that dissipation of the effect largely occurred between Immediate and 1-Day intervals. Dissipation of the effect was confirmed by the lack of significant differences between massed I groups and the combined control groups at either the 1-Day or 7-Day intervals.

Grand mean comparisons of amount of I time showed a significant difference between 1-min. and 5-min. I groups ($t = 2.81, p < .01$); other comparisons were not significant.

Distributed I groups.—The second purpose of the study was to examine the effect on size and persistence of aftereffect, of distributing additions in I time. The groups given 1 min. of massed I time also served as the first point on the variable of distributed I time. The other two points on the variable were represented by groups given 3 or 5 1-min. I periods, each period followed by 2 min. of rest. A simple analysis of variance of the immediately-tested 1-, 3-, and 5-min. distributed I groups showed no difference in amount of initial aftereffect as a function of distributed I time, *F* (2, 42) = .60.

In a two-way analysis of variance of distributed groups over time (0, 1, or 7 days between inspection and test), the interaction term was not significant, *F* (4, 126) = .83. There was no evidence that distributing additional I time had greater effect on the endurance of aftereffect than it did on the size of the immediate effect.

The ineffectiveness of distributing additions in I time was further shown by an analysis of only the 3-min. and 5-min. massed and distributed I groups over the three test periods. There were no significant differences in amount of aftereffect attributable to mode (D) of presentation of the I block (massed or distributed), *F* (1, 168) = .40, to amount (I) of I time (3 or 5 min.), *F* (1, 168) = .21, or to interactions of these variables with test (T) period; D × T groups, *F* (2, 168) = 1.09; I × T groups, *F* (2, 168) = .33; D × I × T groups, *F* (2, 168) = .67. Significant differences emerged as a function of test interval (0, 1, or 7 days between inspection and test), *F* (2, 168) = 8.46, $p < .01$, and the interaction of D × I groups, *F* (1, 168) = 3.36, $p < .05$.

Returning to the two-way analysis of variance, both main variables were significant, I time, *F* (2, 126) = 3.58, $p < .05$, and T interval, *F* (2, 126) = 6.50, $p < .01$. A comparison of Immediate, 1-Day and 7-Day grand means showed that the difference between the Immediate and 1-Day test intervals was significant ($t = 3.66$, $p < .01$) but the other differences were not. The difference between the Immediate and 1-Day means was in the expected direction, and the same direction as found with massed I groups (i.e., a decrease in aftereffect) but the lack of a significant difference between Immediate and 7-Day means was not expected and is a reversal.

Grand mean comparisons of amount of I time showed 3 min. of distributed inspection produced significantly more aftereffect ($t = 2.60$, $p < .01$) than 1 min. Other comparisons were not significant. The fact that 5 min. of massed, but not 5 min. of distributed I time produced significantly more aftereffect than 1 min. accounts for the significant D × I term in the analysis of 3-min. and 5-min. massed and distributed I groups over the three test periods.

All *S*s were given four Posttest trials. Examination of these trials revealed no new information. There was no consistent trend of any kind over the four trials in any group.

DISCUSSION

Suggestions in reviews of studies of visual aftereffects by Duncan (1956), McEwen (1958), and Sagara and Oyama (1957) led to the hypothesis that long amounts of inspection time would have a greater effect on the rate of dissipation of aftereffect than on the size of the immediate effect. In the present study, absence of a significant interaction term in the analysis of amount of inspection time and time of testing does not support the expectation, at least for kinesthetic aftereffect, of a differential decay rate for different amounts of I time. This may represent a functional difference between kinesthetic and visual systems.

Amount of I time over 1 min. did not appear to be an effective variable in the analysis of subgroups ($N = 15$) given 1, 3, and 5 min. of massed or distributed inspection and tested immediately. However, the numerical differences that appeared in the Immediate test groups were maintained across the test intervals (Fig. 1). The result of this pattern was that a significant difference in amount of aftereffect as a function of amount of I time was found when groups were combined across time intervals, providing a larger *N* (45 *S*s vs. 15 *S*s per group). For massed groups, 5-min. I time produced significantly greater aftereffect than 1-min.; for distributed groups 3-min. I time yielded significantly greater effect than 1-min. Thus it appears that magnitude of kinesthetic aftereffect does increase as I time increases beyond 1 min.

A recent study (Bakan, Myers, & Schoonard, 1962) has shown increases in the magnitude of aftereffect with increases in inspection times from 10 to 60 sec. However, measures of the decay function may have been confounded by the effect of repeated testing of *S*s.

The second purpose of the study concerned examination of distributed I time as the possible variable producing the "growth" phenomenon found in kinesthetic aftereffect studies (Charles & Duncan, 1959; Köhler & Dinnerstein, 1947; Wertheimer & Leventhal, 1958). It was hypothesized that in the repeated Inspection-Test, Inspection-Test, etc., procedure, it might be the distribution of I time that created the growing size of the initial aftereffect over time. To test the hypothesis that distributing the inspection periods led to larger effects, groups were given 1, 3, or 5 min. of distributed I time without intervening T periods. There was no evidence from this manipulation that the growth phenomenon could be attributed to distribution of inspection alone, since the analysis of variance of groups tested immediately did not yield a significant difference, nor was there a significant difference between 3 and 5 min. massed inspection as opposed to 3 or 5 min. distributed inspection for subgroups tested immediately (or even for groups combined over test intervals). Therefore, it must be concluded that distributing I time per se does not necessarily lead to a greater initial effect than if the same total amount of I time is given in a single massed period. This suggests that if increasing amounts of aftereffect are found in studies where a test for aftereffect follows each inspection period, the test itself may be contributing to the increasing magnitude of aftereffect.

REFERENCES

BAKAN, O., MYERS, L. B., & SCHOONARD, J. Kinesthetic after-effects and length of inspection period. *Amer. J. Psychol.*, 1962, **75**, 457–461.

CHARLES, J. P., & DUNCAN, C. P. The distance gradient in kinesthetic figural aftereffect. *J. exp. Psychol.*, 1959, **57**, 164–170.

DUNCAN, C. P. On the similarity between reactive inhibition and neural satiation. *Amer. J. Psychol.*, 1956, **69**, 227–235.

KÖHLER, W., & DINNERSTEIN, D. Figural after-effects in kinesthesis. In, *Miscellanea psychologica Albert Michotte.* Louvain: Institut Superior de Philosophie, 1947. Pp. 196–220.

KÖHLER, W., & WALLACH, H. Figural aftereffects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

MCEWEN, P. Figural after-effects. *Brit. J. Psychol. monogr. Suppl.*, 1958, No. 31.

SAGARA, M., & OYAMA, T. Experimental studies on figural aftereffects in Japan. *Psychol. Bull.*, 1957, **54**, 327–338.

WERTHEIMER, M., & LEVENTHAL, C. M. "Permanent" satiation phenomena with kinesthetic figural aftereffects. *J. exp. Psychol.*, 1958, **55**, 255–257.

(Received November 6, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 383–385

END-SPURT IN VIGILANCE[1]

BRUCE O. BERGUM AND DONALD J. LEHR

United States Army Air Defense Human Research Unit, Fort Bliss, Texas

An experiment was performed to test the hypothesis that an end-spurt effect would be demonstrated in a vigilance situation when Ss had knowledge of both length of the vigil and present point in the vigil. 2 groups of 20 Ss each were tested on a visual vigilance task for 135 min. 1 group had knowledge of the length and present point in the vigil and the 2nd group did not. Analysis of the data indicated a statistically significant improvement in performance for the 1st group and no change in performance for the 2nd group in the final period of testing. It was concluded that the experimental hypothesis, based upon an r_g-s_g interpretation of vigilance performance, was supported.

In a discussion of the effects of vigil length on performance, Bergum and Klein (1961) point out that performance generally declines with time, but in some cases (Lindsley, 1944; Wyatt & Langdon, 1932) the effect is confounded by end-spurt improvements in performance. They go on to say that a conditioning interpretation of the effect might imply that one antecedent condition determining the occurrence or nonoccurrence of end-spurt might be previous knowledge of the length of the vigil gained either through practice or through instructions. Under such conditions, the end-spurt, or goal gradient, effect might be described in terms of an r_g-s_g model with termination of the task serving as the reinforcement. A number of testable hypotheses can be derived from this interpretation: (*a*) The end-spurt effect should become more pronounced in successive sessions if the knowledge of vigil length is based upon previous experience. Lindsley (1944) provides supporting evidence for this notion. (*b*) Practice at, or instructions suggesting, one length of vigil and testing at different lengths should yield differently shaped curves. Jerison (1958) provides suggestive evidence in this area. (*c*) Knowledge vs. no knowledge of the length and of the present point in the vigil should yield different results.

The present study was designed to test the latter hypothesis. It was predicted that a group with previous knowledge of the length of vigil and current knowledge of the present point in the vigil would produce an end-spurt effect whereas a comparable group without such information would not.

METHOD

Subjects.—A total of 40 inexperienced National Guard trainees from the Army Training Center, Fort Bliss, volunteered to serve as *S*s in the experiment. These men were between the ages of 18 and 26 yr. and had normal (20/20) vision.

Apparatus.—The apparatus was identical to that employed in earlier studies and described elsewhere (Bergum & Lehr, 1963) and included four test booths, a circular light panel, and an intercom network. The *S* was seated at a small table facing the display which was mounted at eye level on one wall. The 20 lights mounted on the periphery of the circular display illuminated in sequence at a rate of 12 rpm. A signal for this display consisted of the failure of a lamp to illuminate in its normal sequence. Single response pushbuttons were located in each booth and both signals and responses were automatically recorded on paper tape. The signal rate was 12 signals per hour.

---

[1] The research reported in this study was performed for the Human Resources Research Office of the George Washington University under contract to the United States Army. The opinions expressed in this report are solely those of the authors and do not necessarily reflect those of the sponsoring agency.

Conditions.—All *S*s were given a 20-min. pretest on the apparatus, followed by a 20-min. rest and instruction period. The main treatments session ran for a continuous monitoring period of 135 min.

During the main treatment regime the 20 experimental *S*s were given information regarding the exact length of the vigil. This group was allowed to wear, or was provided with, timepieces and was encouraged to make use of them during the session. The remaining 20 control *S*s received no special treatment, were given no knowledge of vigil length, and were required to deposit all timepieces with *E*s before entering the booths.

RESULTS

A comparison of the experimental and control groups in terms of pretest performance indicated identical means for the two groups and no statistical tests were performed.

Because it was hypothesized that the experimental condition would lead to a significant improvement in detection performance during the final period of testing, the analyses were confined to comparisons between the groups in terms of performance changes during this final period. For the experimental group, the results indicated an 18% increase in detections between the final and preceding time periods. For the control group, a 1% increase was observed.

The significance of the observed effect was tested in two ways. In the first case, a simple analysis of variance, reported in Table 1, was performed on the difference scores for individuals between the fourth and the final test periods for both groups, $F\ (1,\ 38) = 4.38$, $p < .05$.

TABLE 1

ANALYSIS OF VARIANCE OF CORRECT DETECTIONS (Arc-Sine Transformation)

Source	*SS*	*df*	*MS*	*F*
Between	11.02	1	11.02	4.38*
Within	98.75	38	2.60	
Total	109.77	39		

$^*p < .05$.

In the second analysis direct-difference *t*'s were computed for the two groups in terms of detection score differences between the fourth and final periods. The *t* for the control group was less than 1, indicating no change in performance between the final two time periods. The *t* for the experimental group, however, was significant beyond the .01 level, $t(19) = 4.18$. These results indicate a significant improvement in performance in the final testing period as a result of the experimental condition. The hypothesis that knowledge of vigil length will lead to a significant end-spurt effect in a vigilance task was supported by the results.

DISCUSSION

The present results, in combination with those of Jerison (1958) and Lindsley (1944), appear to be consistent with the suggested r_g-s_g model of vigilance performance. It should be noted that Jerison made an expectancy interpretation of his own results. The expectancy model, however, would not appear to apply to the Lindsley results nor the results of the present study without further assigning motivational properties to the expectancy. Because of the self-stimulating characteristic of goal-oriented r_g-s_g's (Spence, 1956), increases in the frequency and intensity of fractional anticipatory goal responses as termination of the task is approached would result in a general increase in the motivational level of the monitor with a consequent increase in the probability of detection. Such an interpretation is in line with an activation interpretation, also, and provides a mechanism to account for the increases in general stimulation level required by such an interpretation.

REFERENCES

Bergum, B., & Klein, I. C. A survey and analysis of vigilance research. *HumRRO res. Rep.*, 1961, No. 8.

Bergum, B., & Lehr, D. J. Effects of authoritarianism on vigilance performance. *J. appl. Psychol.*, 1963, **47**, 75–77.

Jerison, H. J. Experiments on vigilance: Duration of vigil and the decrement function. *USAF WADC tech. Rep.*, 1958, No. 58–369.

Lindsley, D. B. Radar operator "fatigue": The effects of length and repetition of operating periods on efficiency of performance. *USAF OSRD Rep.*, 1944, No. 6.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Wyatt, S., & Langdon, V. N. *Inspection processes in industry.* London: His Majesty's Stationery Office, 1932.

(Received November 13, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 386–393

MONOPTIC AND DICHOPTIC VISUAL MASKING[1]

PETER H. SCHILLER[2] AND MORTON WIENER

Clark University

This study investigated the extent of visual masking (metacontrast), under both monoptic and dichoptic conditions, of relatively complex visual forms (capital letters). Comparisons were made between the masking effects of homogeneous fields of varying reflectances and patterned fields of various degrees of patterning. The results show that masking varied both with reflectance and patterning, but in a different manner. Homogeneous fields, ranging from white through intermediate grays to black, produced more masking the less the reflectance. By contrast, the patterned fields did not produce a corresponding monotonic increase with increasing density of pattern; masking was minimal with low and high density and maximal with intermediate density of pattern. The extent of the masking at these intermediate points was much higher than the masking with any of the homogeneous fields. Masking was obtained for both monoptic and dichoptic presentations. However, the extent of the masking decreased with practice on dichoptic presentation, when the masking fields were patterned; monoptic presentation did not yield such decrement.

The inhibition of the first of two successive visual stimuli falling upon adjacent retinal areas was observed as early as 1868 by Exner, and was investigated later by Helmholtz and Baxt (Baxt, 1871), Sherrington (1897), McDougall (1904), Stigler (1910), Baroncz (1911), Fry (1934), and Piéron (1935). These investigators employed primarily flashes of light in the study of this phenomenon which was named metacontrast by Stigler (1910).

The experiments of Werner (1935, 1940) are the first to systematically explore this phenomenon with visual forms. His studies were directed toward the investigation of contour formation. Werner used a broad array of stimuli in the visual masking situation, the best known of which is the disk-ring sequence. The results of these studies indicated that the formation of contours takes time; when this process is hampered by other stimuli of specific size, shape, and time relation to the first stimulus, the contours of the first figure will not become established, and thus the figure will not be perceived, or will be perceived only partially.

Recently a number of *E*s have investigated the visual masking phenomenon (Alpern, 1952, 1953; Fehrer & Raab, 1962; Kolers & Rosner, 1960; Lindsley, 1961). Several important variables were studied systematically, showing that the contour relation between the figures, their relative exposure time, the time interval between the presentation of the first and second figure, the rate of recycling, and the retinal locus of their presentation are important variables in the extent of visual masking.

Experimental evidence on visual masking with dichoptic presentation, i.e., when the first stimulus is presented to one eye and the second stimulus to the other eye, is unclear and in some cases contradictory. Stigler (1910, 1913, 1914, 1926) was, for the most part, unsuccessful in his attempts to obtain masking with dichoptic presentation. Werner (1940), using the disk-ring sequence presented at corresponding retinal points in the two eyes, found masking effects comparable to monoptic conditions, but found, as did Baumgardt and Segal (1942), that the time constants under the two presentation conditions were different. Kolers and Rosner (1960), also using the disk-ring sequence, found good masking dichoptically, although the time course involved in their study was different from that found by Werner. Kolers and Rosner also obtained masking on dichoptic presentation with more complex figures. On the other hand Alpern (1953) was unable to obtain any masking with dichoptic presentations. Furthermore, Alpern could not obtain any metacontrast when the stimuli were made to fall on the fovea, although Werner reported earlier that masking does occur under such conditions, though not as readily as with nonfoveal presentations. Recently Lindsley (1961) found that a bright flash following a figure masked only when both stimuli were presented to the same eye.

---

[1] This research was supported by a National Institute of Mental Health grant (M-3860) to the second author and by a Mental Health postdoctoral fellowship (FPD-11, 589-C1) to the first author. The authors are indebted to Hans-Lukas Teuber for his helpful suggestions.

[2] Now at the Massachusetts Institute of Technology, Cambridge.

In light of these findings it becomes important to further assess the extent of visual masking under both monoptic and dichoptic conditions, using visual forms with relatively complex contours, where the extent of the figural patterning of the masking field is systematically varied. Our study therefore attempted to investigate three questions: (*a*) What is the extent of visual masking of relatively complex stimuli when the degree of figural patterning of the masking field is systematically varied? (*b*) What is the extent of masking when the luminance of a nonpatterned masking field is varied? (*c*) How is the extent of masking affected by monoptic and dichoptic presentations?

METHOD

Subjects.—Four paid volunteer *S*s who had equal visual acuity in both eyes were used (20/20 vision for all *S*s, one corrected with

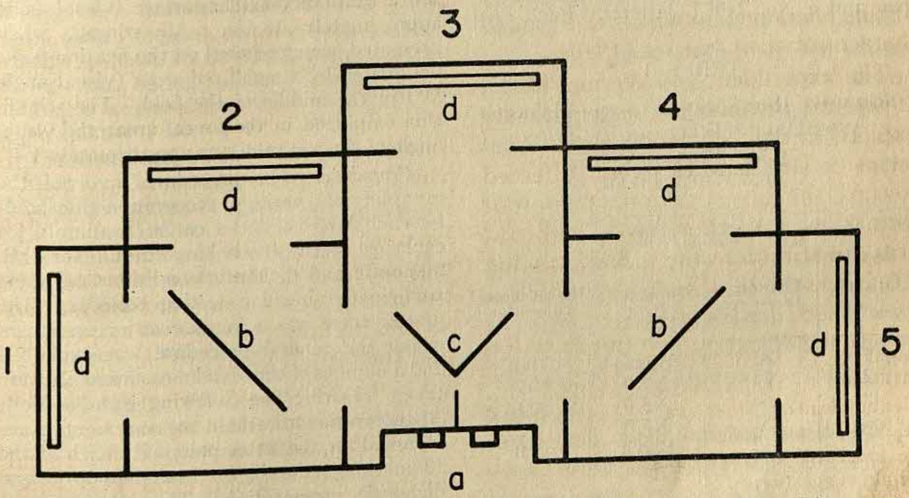


FIG. 1. Top view of stereoscopic viewer, showing the five fields: (a) eyepiece; (b) two-way mirrors with 50% transmitting and 50% reflecting characteristics; (c) two-way mirrors with 33% transmitting and 67% reflecting characteristics; (d) tachistoscopic tubes, two per field.

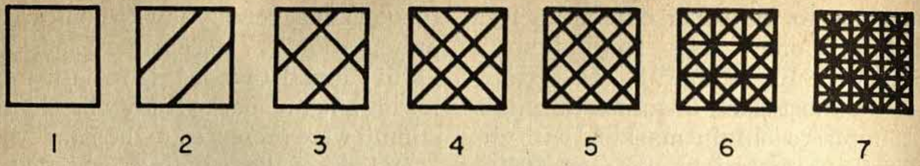


FIG. 2. Stimulus figures of the patterned set.

glasses). These *S*s had served in an earlier experiment (Schiller & Wiener, 1962) in which the same equipment was used; they were thus familiar with the viewing characteristics of the apparatus. However, *S*s were naive about the conditions and purposes of this study.

Apparatus.—The apparatus was a five-field tachistoscopic viewing box driven by a tachistoscopic timer (Scientific Prototype Manufacturing Company, Modified Model 1B). In this apparatus stimuli could be presented both simultaneously or in succession to either of the eyes. The duration of exposure, the interval between two successive fields, and the luminance of the fields could be varied independently in each field. (A schematic drawing of the viewing box is shown in Fig. 1.) The light sources were cold cathode mercury vapor argon lamps, coated with magnesium and tungstate phosphor, with a maximum rise and fall time of 2½ μsec. (purchased from Scientific Prototype Manufacturing Company).

Materials.—Identification stimuli ("A" stimuli): 60 5 × 7 in. cards, each having two capital letters drawn on it in black India ink with a Leroy drawing set, using a No. 3 pen and a No. 240CL letterer. Each card contained a different combination of two letters. The letters I, J, M, and W were omitted, as were all combinations forming a word.

TABLE 1

RELATIVE REFLECTANCE OF THE MASKED AND MASKING FIELDS IN RELATION TO THE REFLECTANCE OF A WHITE FIELD

Masking Field			
Stimulus Card	Pattern Dimension (%)	Stimulus Card	Homogeneous Dimension (%)
1	85	White	(100)
2	74	N7.0	47
3	66	N5.0	11
4	51	N3.0	2
5	42	Black	0
6	27		
7	7		

Masked Field: 87%–90%

Masking stimuli ("B" stimuli): Two sets of masking stimuli were prepared: (*a*) Patterned set: This set is shown in Fig. 2. These stimuli were drawn on 5 × 7 in. cards employing the same drawing pen and ink as above. The squares measured 16 × 16 mm. (*b*) Homogeneous set: This set, also on 5 × 7 in. cards, consisted of a homogeneous white field (blank card), three homogeneous shades of gray squares, No. N7.0/6754, N5.0/6835, and N3.0/6611 of the Munsell series, and a black square. These stimuli, with the exception of the blank card, also measured 16 × 16 mm. The blank card and the black square of the second set can be considered to be the extreme points of the first set also. For the reflectance value of the A and B stimuli see Table 1. The percent values in this table represent the reflectance of the stimuli in relation to the reflectance of a white field of the same dimensions.

The *S*s looking into the viewing apparatus saw a white rectangular surface (visual angle approximately 18.4°). All stimuli, when presented, were centered on the fixation point designated by a small red cross (visual angle .3°) in the middle of the field. The stimuli thus impinged in the foveal area; the visual angle of these stimuli was approximately 1.4°.

Procedure.—The procedure involved the obtaining of a series of recognition thresholds for each *S* on the letter cards (A stimuli) for each one of the B masking stimuli, for both monoptic and dichoptic conditions. The *S*s participated in nine testing sessions. The first of these was a familiarization session in which the general procedure was explained and a number of threshold measurements were taken. During the following eight sessions 20 recognition threshold measurements were obtained for the letter pairs with each of the 12 masking stimuli, for both monoptic and dichoptic presentation. Thus there was a total of 480 threshold measurements for each *S*.

During every session threshold measurements were obtained for five letter pairs with each of the 12 masking stimuli, using one of the visual viewing conditions. The order of the letter pairs and the 12 masking fields was randomized for each session.

The monoptic and dichoptic sessions were arranged as follows:

Testing Days	*S*s 1 and 2	*S*s 3 and 4
1 & 2	Monoptic	Dichoptic
3 & 4	Dichoptic	Monoptic
5 & 6	Monoptic	Dichoptic
7 & 8	Dichoptic	Monoptic

Each testing session was preceded by a 10-min. adaptation period, during which *S* was seated near the apparatus in a semilit room and was asked to relax. Following this period four threshold measurements were taken for warm up, using two-letter words such as IT, AM, DO, etc.

Thresholds for the A stimuli were measured by the method of ascending limits, starting well below the threshold obtained during the warm-up period. Increments were presented in .5-msec. steps. The *S*s were instructed to avoid wild guessing, but were asked to make their best guesses.

There was a 3–4 sec. interval between successive exposures of the recognition card (A stimuli). The *E* alerted *S* by saying, "ready" before each exposure. There was a 10–15 sec. interval each time the A stimulus was changed, and a 20–25 sec. interval when both the A and B stimuli were replaced (always after 5 threshold measurements). During these intervals *S* was told to relax. With these frequent rest intervals it was possible to make 60 threshold measurements during the testing sessions without any noticeable fatigue effects. Each session lasted approximately 1 hr.

The identification stimuli always appeared in Field 4. The masking stimuli appeared in Field 5 for monoptic presentation and in Field 2 for dichoptic presentation. The interstimulus interval between the A and B fields was always 40 msec., during which interval Field 3 came on again. This interval was chosen on the basis of the study by Kolers and Rosner (1960) in which the greatest masking was found to occur between intervals of 40–80 msec. Pilot work showed good masking at the 40-msec. interstimulus interval. The B masking stimuli always appeared for 100 msec. Only the duration of the A stimuli was varied for the threshold measurements.

The illuminance of each of the fields without the stimulus cards, as measured by a MacBeth illuminometer, gave a mean reading of 1.6 ft-c at the eyepiece.

RESULTS

The mean recognition threshold was the duration of the exposure of the masked (A) stimuli in milliseconds, at which *S*s correctly recognized a given letter pair. Since examination of the data showed no significant *S*s × Conditions interaction, only combined data for all *S*s are presented here.

The mean recognition thresholds and *SD*s for the pattern and homogeneous dimensions are shown in Table 2 for both the monoptic and dichoptic conditions of presentation. Thresholds for the "white" field represent, essentially, the point at which *S*s can correctly identify letter pairs without masking, since the reflectance of this field is identical with that of the fixation field. Under this condition *S*s report only the brief appearance of letter pairs on a white surface.

The data for the homogeneous dimension show a gradually increasing monotonic function from "white" to "black." This increase is very small. However, t tests comparing the white and black masking field data show that the latter produced significantly higher thresholds for all four *S*s ($t = 5.72$, 5.00, 2.78, 2.48 for each of the four *S*s).

As can be seen the extent of the visual masking for the pattern dimension is highest with the No. 3, 4, 5, and 6 masking stimuli, being less at both ends. The masking is more pronounced monoptically than dichoptically.

The data were further analyzed to assess more carefully the difference between monoptic and dichoptic presentations for the pattern dimension. This analysis was carried out on the

TABLE 2

MEAN RECOGNITION THRESHOLDS OF MASKED STIMULI IN MILLISECONDS WITH EACH MASKING STIMULUS

	Pattern Dimension								
	White[a]	1	2	3	4	5	6	7	Black[a]
Monoptic									
M	7.0	8.0	10.1	13.6	13.5	12.7	12.9	9.9	8.6
SD	1.0	1.1	1.8	3.6	4.5	3.3	3.5	1.7	1.5
Dichoptic									
M	6.4	7.5	10.4	10.8	10.7	10.1	10.4	8.5	7.4
SD	1.4	1.5	4.1	2.5	3.5	2.7	3.6	1.7	1.5

	Homogeneous Dimension				
	White[a] (100% reflectance)	N7.0 (47% reflectance)	N5.0 (11% reflectance)	N3.0 (2% reflectance)	Black[a] (0% reflectance)
Monoptic					
M	7.0	7.5	8.4	8.4	8.6
SD	1.0	1.4	1.1	1.5	1.5
Dichoptic					
M	6.4	6.8	7.1	7.1	7.4
SD	1.4	1.3	1.5	1.4	1.5

[a] The data for the "white" and "black" B fields are presented twice, appearing both under the Pattern and Homogeneous dimensions for easier comparison.

basis of *E*'s observations during the course of the experimental sessions. The data in Table 2 show that the extent of masking is greater with monoptic than with dichoptic presentation. This suggests that part of the masking effect may originate in the retina. However, further inspection of the data shows that the reduction in the masking effect may be attributed to other factors. Figure 3 shows a reanalysis of the pattern dimension, in which the scores were divided into the first and second set of 10 measurements obtained for each *S* under each condition. These graphs show that the differences between monoptic and dichoptic presentation are very small for the first set of 10 threshold measurements, but differ markedly for the second set of 10 measurements. These findings suggest that practice affected monoptic and dichoptic masking differently. It should be noted that the abscissa in this graph is laid out in arbitrary units, since neither the degrees of pattern, as shown in Fig. 2, nor the reflectances, as shown in Table 1, were scaled to yield equal intervals.

The data show a higher variance in the middle region of the pattern dimension than at the ends. This implies that masking in the middle region was more effective for some of the letter pairs than others.

DISCUSSION

The results of this experiment show the following: (*a*) There is a marked increase in the recognition threshold for letter pairs with those masking stimuli which fall in the middle of the pattern dimension. The thresholds are lower at both extremes. (*b*) There is a small monotonic increase in recognition threshold with the masking stimuli in the homogeneous dimension, going from high to low reflectance. In comparison with the pattern dimension this masking is very slight. (*c*) The magnitude of masking under monoptic and dichoptic conditions is nearly identical for the first set of 10 measurements in the pattern dimension. For the second set of 10 measurements the monoptic condition yields about the same magnitude of masking as for the first 10 measurements, but in the dichoptic condition there is a marked drop in threshold, particularly in the middle range. (*d*) There is an increase in variance in the middle region of the pattern dimension which suggests the possibility that specific masking is involved.

The fact that the specificity of the contour relationship of the A and B stimuli is important has been shown by Werner (1935) and Kolers and Rosner (1960). These investigators found a decrease in the masking effect as the separation between the contours of the masked and masking fields was increased. In our study very little masking was found in the homogeneous dimension, where there were no common contours between the A and B stimuli. This observation, as well as the increase in variance in the pattern dimension, may be taken as further evidence for the view that specific contour relation between the two stimuli is critical in bringing about extensive masking.

It is difficult to account, with a single explanatory concept, for both the greater masking and the increased variability in the middle portion of the pattern series. Although, as Werner, and Kolers and Rosner have pointed out, the degree of common contour overlap in the masking and masked stimuli is an important variable, in our study were this the only significant parameter, the maximum masking should have occurred for Masking Stimulus No. 7. With this stimulus there is the greatest probability of common contour overlap, if only because

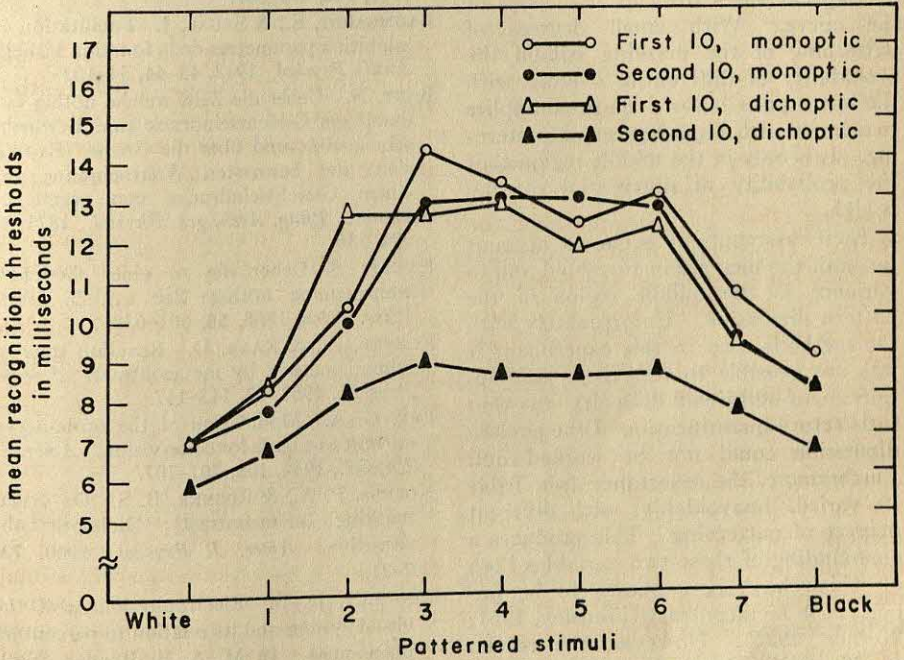


FIG. 3. Mean recognition thresholds for the first and second set of 10 measurements taken for each patterned field under monoptic and dichoptic presentation conditions. (The points on the abscissa are not quantitatively ordered—see text.)

there are more lines in this pattern. It would seem thus that some other parameter or parameters must be specified. One possibility is that greatest masking in our situation depends on two criteria: (*a*) the relative similarity of the pattern contrast of the masked and masking stimuli, and (*b*) the maximal physical overlap of these stimuli. The high recognition thresholds in the middle region can be accounted for by these assumptions. This explanation implies an imbedding hypothesis; maximal imbedding here is assumed to take place when the masked and masking stimuli are structurally similar. The increased variance in the middle region can be explained by assuming that the extent of common contour overlap differed for each of the masked (e.g., an L vs. V) stimulus pairing. When there was much overlap, the threshold was relatively high, and when there was little overlap, the threshold was relatively low. Due to the configuration of the masking stimuli the variability of specific contour overlap decreases towards both ends of the curve. With small degrees of patterning of the masking stimuli the probability of differential overlap with the letter pairs is low; the same applies to stimuli with high degrees of patterning. It is only in the middle region that the probability of differential overlap is high.

These assumptions seem to account for both the increase in threshold and in variance in the middle region of the pattern dimension. Unfortunately with the methods used in this experiment it was not possible to test these assumptions. An additional difficulty was that satisfactory quantification of the pattern dimension could not be worked out. Furthermore, the reflectance (see Table 1) varied, unavoidably, with different degrees of patterning. This produces a confounding of these two variables both of which have been shown to be important in metacontrast (Lindsley, 1961; Werner, 1935). However, the results of this study show that reflectance in itself (homogeneous dimension) affects masking in a manner different from patterned configurations. Studies further investigating these problems are presently under way.

The finding that masking decreased with increasing number of trials, but only under dichoptic and not monoptic conditions, suggests some kind of perceptual "accomodation," which may involve suppression of one eye, or some process related to binocular interaction. Since *S*s were not aware of the dichoptic presentation of the stimuli, this question becomes especially intriguing. A series of experiments are under way to assess the possible mechanisms of this practice effect of repeated dichoptic presentation.

REFERENCES

ALPERN, M. Metacontrast: Historical introduction. *Amer. J. Optom.*, 1952, **29,** 631–646.

ALPERN, M. Metacontrast. *J. Opt. Soc. Amer.*, 1953, **43,** 648–657.

BARONCZ, Z. Versuche über den sogenannten Metakontrast. *Pflüg. Arch. ges. Physiol.*, 1911, **140,** 491–507.

BAUMGARDT, E., & SEGAL, J. Facilitation et inhibition parametres de la fonction visuell. *Année Psychol.*, 1942, **43–44,** 54–102.

BAXT, N. Ueber die Zeit, welche nöthig ist, damit ein Gesichtseindruck zum Bewusstsein kommt und über die Grosse (Expansion) der bewussten Wahrnehmung bei einem Gesichtseindrucke von gegebener Dauer. *Pflüg. Arch. ges. Physiol.*, 1871, **4,** 324–336.

EXNER, S. Ueber die zu einer Gesichtswahrnemung nöthige Zeit. *S. B. Akad. Wiss. Wien*, 1868, **58**, 601–632.

FEHRER, E., & RAAB, D. Reaction time to stimuli masked by metacontrast. *J. exp. Psychol.*, 1962, **63,** 143–147.

FRY, G. A. Modulation of the optic-nerve current as a basis for color-vision. *Amer. J. Physiol.*, 1934, **108,** 701–707.

KOLERS, P. A., & ROSNER, B. S. On visual masking (metacontrast): Dichoptic observation. *Amer. J. Psychol.*, 1960, **73,** 2–21.

LINDSLEY, D. B. Electrophysiology of the visual system and its relation to perceptual phenomena. In M. A. B. Brazier (Ed.), *Brain and behavior.* Vol. 1. Washington, D. C.: American Institute of Biological Sciences, 1961. Pp. 359–392.

McDougall, W. The sensation excited by a single momentary stimulation of the eye. *Brit. J. Psychol.*, 1904, **1**, 78–113.

Piéron, H. Le processus de métacontraste. *J. Psychol. norm. pathol.*, 1935, **32**, 5–24.

Schiller, P. H., & Wiener, M. Binocular and stereoscopic viewing of geometric illusions. *Percept. mot. Skills*, 1962, **15**, 739–747.

Sherrington, C. S. On reciprocal action in the retina as studied by means of rotating discs. *J. Physiol.*, 1897, **21**(1), 33–54.

Stigler, R. Chronophotische Studien über den Umgebungskontrast. *Pflüg. Arch. ges. Physiol.*, 1910, **134**, 365–435.

Stigler, R. Metacontrast (Demonstration). *Arch. int. Physiol.*, 1913, **14**, 78.

Stigler, R. Demonstration des Metakontrastes mit Hilfe des Metakontrastapparates. *Zbl. Physiol.*, 1914, **29**, 194–196.

Stigler, R. Die Untersuchung des zeitlichen Verlaufes der optischen Erregungen mittels des Metakontrastes. In E. Aberhalden (Ed.), *Handbuch der biologischen Arbeitsmethoden.* Berlin: Verlag von Urban Schwarzenberg, 1926. Pp. 949–968.

Werner, H. Studies on contour: I. Qualitative analysis. *Amer. J. Psychol.*, 1935, **47**, 40–64.

Werner, H. Studies on contour strobostereoscopic phenomena. *Amer. J. Psychol.*, 1940, **53**, 418–422.

(Received November 13, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 394–398

EFFECTS OF KNOWLEDGE OF RESULTS ON A VERBAL MEDIATING RESPONSE[1]

MARGARET JEAN PETERSON

Indiana University

Knowledge of results (KR) was given following 0, 2, 4, or 36 sets to determine the effect upon mediated responses and response availability when single verbal mediating chains were established. Using CCCs, 1 pair (A-B) was shown followed by a 2nd pair with the 1st stage response as the stimulus (B-C). In the test stage, the 1st stage stimulus was presented with the 2nd stage response and 2 other CCCs (A-CDE). S selected the best response to pair with A. In the control condition, an extraneous syllable was substituted for B. The frequencies of presentation of the first 2 stages were 1, 3, or 6. Although both mediated responding and response availability were clearly demonstrated, neither was systematically influenced by KR or by the frequency of presentation.

Mediated behavior has been demonstrated when no reasonably direct form of reinforcement could be identified (e.g., Bugelski & Scharlock, 1952; Peterson & Blattner, 1963). In the latter study, a single mediating chain was established by presenting one pair (A-B), then a second pair whose stimulus was the response member of the previous pair (B-C), and testing by a multiple-choice type arrangement in which the stimulus from the first pair was presented with the response from the second pair and two additional alternatives (A-CDE). The *S* was to select the alternative which made the "best pair" with A. An extraneous syllable was substituted for B in one of the first two stages for control presentation. The *S*s read off all materials and no learning trials in the usual sense were included. Thus no specific reinforcement other than auditory feedback was given during the experimental procedure; however, evidence of mediated responding and of a general facilitating effect (response availability) was indicated.

In the present studies, it was assumed that given information about the correctness of the test trial selection, the entire experimental situation might become more like a problem in concept formation for *S*. If he solved the mediational paradigm, future responses would reflect more conscious associations, thus facilitating mediated responding and depressing response availability. If the influence of reinforcement were minimal, knowledge of results (KR) should have no differential effect upon indexes of either mediation or response availability.

The frequency of the first two stages was also varied. Earlier studies (Peterson & Blattner, 1963) showed that response availability reflected increased frequency, but mediation selections did not increase more than would have been expected on the basis of response availability.

METHOD

The construction of learning materials has been described in detail elsewhere (Peterson & Blattner, 1963). CCCs of 0–17% associa-

---

[1] This research was supported by Grant M-5209 from the National Institute of Mental Health. The assistance of Keith Blattner is gratefully acknowledged.

tion value (Witmer, 1935) were used to construct 36 sets consisting of a single A-B pair, a B-C pair, and an A-CDE test trial. On the test trial the stimulus of the first stage, A, was presented with the response of the second, C, and two other response alternatives drawn from CCCs that served as C syllables on other sets. The appropriate C syllable appeared equally often as the first, second, or third alternative. Control sets were constructed by substituting an extraneous syllable for C in the first stage (Control I) or in the second stage (Control II). Test trials were identical for a set under mediation and the two control presentations.

Frequency of presentation was manipulated by presenting the first stage one, three, or six times, followed by the second stage shown with the same frequency as the first.

Four treatments differed only in the number of sets followed by knowledge of the correctness of the test trial selection (KR). The materials were identical for all four. In 0KR, no information was given so the procedure was the same as previously reported (Peterson & Blattner, 1963, Exp. I). For the other three conditions KR was given following the first two sets (2KR); following the first four sets (4KR); or following all 36 sets (36KR). In the 2 and 4 KR conditions, the sets followed by KR were equally divided between mediation and control presentation.

The design was 4 (Number of Sets Followed by KR) × 3 (Type of Presentation) × 3 (Frequency of Presentation). Each set appeared equally often in the resulting 36 cells over all *S*s.

The *S*s were instructed:

> In this experiment you will be shown some consonant syllables. A consonant syllable is composed of three consonant letters. When the shutters are opened you will see a pair of consonant syllables. You are to read off the letters aloud. Then the drum will turn and you will see more pairs of consonant syllables. You read off the letters every time the drum turns. Soon you will see another kind of presentation. Four consonant syllables will be shown. The one on your left is the stimulus syllable and you are to read it off with each of the other syllables, the response syllables. For example, if ABC DFG appeared you would say, "ABC DFG." Then reread the stimulus syllable, ABC, with the next response syllable, ABC HJK, and then reread the stimulus syllable with the last syllable. You are to decide which of the last three syllables seems "right" or makes the best pair with the stimulus syllable. Do you have any questions? Let's try a few. The practice trials are to acquaint you with the general procedure.

If necessary, the instructions were reread or paraphrased.

For 2 and 4 KR conditions the instructions continued,

> When we begin the main part of the experiment, I will tell you if your choice was right or wrong on the first two (four) trials. Then I won't say anything following the rest of the trials. There is a system so that you could be right on all trials.

The modification for 36KR read,

> When we begin the main part of the experiment I will tell you if your choice was right or wrong for all the trials. There is a system so that you could be right on all trials.

Three practice sets which differed from any of the experimental types of presentation by having a new syllable substituted for A on the test trial were shown before the 36 sets. A practice set demonstrated each of the three frequencies of presentation.

Procedure and apparatus.—All materials were shown on a memory drum at a 4-sec. rate. The drum was stopped on the test trials. After the experiment, *S*s were questioned and dismissed.

Subjects.—The *S*s were 96 students from introductory psychology courses at Indiana University who participated to fulfill a course requirement. There were 24 *S*s assigned to each of the four KR conditions.

RESULTS

As Table 1 shows, increasing the number of sets followed by KR did not reliably affect responses for any of the types of presentation nor for the different frequency of presentation conditions. The *F* for the between effect of Number of KR sets was less than unity. Reliably more C selections were made following mediational presentation than either of the controls, giving evidence that mediational chains were established. The *F* for the third main effect, Frequency

TABLE 1

PROPORTIONS OF C SELECTIONS ON THE TEST TRIALS

Type of Presentation	Frequency of Presentation	Number of Sets Followed by KR			
		0	2	4	36
Mediation	1	.65	.60	.67	.53
	3	.59	.61	.59	.55
	6	.67	.52	.58	.61
Overall		.64	.58	.61	.57
Control I	1	.39	.55	.55	.57
	3	.48	.52	.52	.52
	6	.56	.55	.63	.53
Overall		.48	.54	.57	.54
Control II	1	.46	.52	.57	.53
	3	.43	.51	.46	.56
	6	.46	.61	.58	.54
Overall		.45	.55	.54	.55

of Presentation, was nonsignificant, F (2, 184) = 2.39.

Although more mediation and fewer control responses were given during the 0KR condition, the Type of Presentation × Number of KR Sets interaction term was not statistically significant. Because a sizable increase in control responding was observed for the KR conditions, a t test was computed between the control presentations of 0KR and 36KR, the largest difference. The value obtained, 1.82 with $df = 94$, provided additional corroboration of the statistical nonsignificance of the trend.

If selection of the C alternative on the test trials was unaffected by *S*s having experienced it during the second stage, its probability of occurrence would be .33. All of the proportions shown in Table 1 for the control conditions exceeded the chance level. The t for the difference of the lowest proportion, .45 for Control II, 0KR, was 2.69, significant between the .02 and the .01 levels with $df = 23$. Response availability was clearly reflected.

It was possible the learning manifested during the last sets would be camouflaged by earlier errors. The data for the 0KR and for the 36KR experiments were analyzed into the four presentations of each of the combinations. If KR influenced the results, an increase from the first to the fourth presentation would be expected for mediation presentation and a decrease for control presentation; however, no reliable evidence was found to support such expectations

TABLE 2

ANALYSIS OF VARIANCE OF C SELECTIONS

Source	*df*	*MS*	*F*
Between *S*s			
*S*s	95	4.1247	
No. sets with KR (KR)	3	1.6373	—
Error (b)	92	4.2058	
Within *S*s			
Type of presentation (T)	2	8.6076	7.49***
Frequency of presentation (F)	2	2.0868	2.39
T × F	4	.5226	—
T × KR	6	2.0182	1.76
F × KR	6	.6316	—
T × F × KR	12	1.0071	1.52
Error (1)	184	1.1492	
Error (2)	184	.8740	
Error (3)	368	.6641	
Total	863		

*** $p < .001$.

when Cochran *Q* tests for related samples were applied.

The analysis of *S*s' responses to the postexperimental interrogation showed that no *S* was able to outline the mediation paradigm nor did any indicate awareness that the entire sequence of stages should be considered. The reports included the use of sound rhythms, "what sounded best," alphabetical sequences, idiosyncratic associations, and random guessing. The *S*s invariably concluded with the statement that they did not know the system used to identify right or wrong selections.

DISCUSSION

Apparently, providing information about the accuracy of the choice on the test trial did not have a systematic effect upon either mediated responses or response availability even when KR was given on all 36 sets. KR did not appear to interact with any other variable; no evidence was found for a practice or learning effect over the experimental trials. Thus no factors were noted which might have obscured or confounded the influence of KR.

The ineffectiveness of KR on all measures was curious because of the relative simplicity of the design albeit disguised by presentation as CCC pairs. The *S*s participating under KR conditions were told that a system existed which could be used to select "correct" answers. C selections were identified as being correct only following mediation presentation; D and E responses were considered correct on control sets so *S* had ample opportunity to distinguish between the two types of presentation, particularly under the 36KR condition, and yet not one *S* reported awareness of the experimental design.

Generally, KR or reinforcement in some form has been a potent variable. One clue to its ineffectiveness upon the establishment of a mediating chain may be found in the postexperimental reports of *S*s. Despite the inability to verbalize a solution, a differential effect favoring mediation presentation was systematically observed. The regularity with which *S*s claimed to have made choices based on what "sounded right" suggested a sort of "greasing of the sound track skids" with mediation presentation, and the absence of awareness of the experimental design indicated that the establishment of a mediational chain was not a conscious process. Other investigators have come to the same conclusion (e.g., Bugelski & Scharlock, 1952; Peters, 1935). If the process were not a conscious one, *S* would not be able to profitably apply knowledge gained from *E*'s reinforcements. Perhaps the establishment of a mediating chain, probably a central process, simply is not amenable to some of the variables which have been demonstrated to be important when applied to other psychological phenomena.

However, the auditory effects from verbalization of the pairs may be considered one form of reinforcement as implied by Birge's (1941) study. She found that children gave evidence of mediated generalization only when they verbalized the mediating labels. And yet, Kendler, Kendler, and Wells (1960) indicated that mere posession of a verbal label that could serve as a mediator did not necessarily facilitate learning. They report evidence that reversal shifts increased with increasing age although the incidence of nonreversal shifts did not decrease as might be expected. Perhaps a certain level of maturation is required for the establishment of mediating responses.

Although both mediation and response availability were observed under the four conditions of KR, neither increased reliably with additional presentations of the acquisition stages. No statistical analysis of the individual KR conditions was made because the main effect and all interactions involving Frequency were nonsignificant. The data from the 0KR condition was compared with a similar experiment (Exp. I) in an earlier study (Peterson & Blattner,

1963) in which frequency of presentation of the acquisition stages was found to be an influential variable, and no reliable differences were found, indicating that frequency was effective with the present 0KR condition. Inspection of Table 1 suggested that the general nonsignificance over the four conditions reported herein probably arose from the variable and nonsystematic relationships with the 2, 4, and 36KR conditions. Motivational or other aspects associated with KR may have been sufficiently disrupting to obscure the operation of the Frequency variable.

Had a different technique been used for presentation of the repetition conditions: e.g., A-B, B-C, A-B, B-C, etc., a more vivid effect might have accrued to frequency of presentation. Such a technique would also tend to provide additional KR as would insertion of cue trials consisting of the stimulus alone from the acquisition pair presented immediately prior to the cue trial. Errors made by *S* in providing the correct response could be used as an index of the strength of the association prior to the test trial. While other procedures could have been used to vary the frequency of presentation of the acquisition stages, the technique employed herein could not be said to exert a systematic effect upon either mediating responses or upon response availability.

REFERENCES

Birge, J. S. The role of verbal responses in transfer. Unpublished doctoral dissertation, Yale University, 1941.

Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.

Kendler, T. S., Kendler, H. H., & Wells, D. Reversal and nonreversal shifts in nursery school children. *J. comp. physiol. Psychol.*, 1960, **53**, 83–88.

Peters, H. Mediate association. *J. exp. Psychol.*, 1935, **8**, 20–48.

Peterson, M. J., & Blattner, K. C. Development of a verbal mediator. *J. exp. Psychol.*, 1963, **66**, 72–77.

Witmer, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–360.

(Received November 14, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 399–408

FREE RECALL OF REDUNDANT STRINGS OF SYMBOLS BY CHILDREN[1]

RICHARD S. BOGARTZ AND EDWARD C. CARTERETTE

University of Iowa *University of California, Los Angeles*

The learning of sequentially constrained strings of symbols was studied by requiring 192 4th-, 5th-, and 6th-grade children to recall freely strings of letters or digits. Each S learned first a highly constrained (L) or random (R) list, followed by either an L or an R list. Conclusions were: (a) L strings were learned faster than R strings; (b) digits were learned faster than letters, and (c) the higher S's grade was the faster either type of string was learned; (d) LL and RR groups performed equally well whether the 2nd list was new or identical with the 1st; (e) LR and RR groups did not differ, but (f) RL groups did significantly worse than LL groups; (g) compared with adults, children learned constrained strings very slowly; (h) however, the data contained marked evidence that certain constraints were learned well and quickly.

Shannon (1948) in his development of selective information theory first estimated the sequential redundancy of standard printed English as being about 50%, using a behavioral method. That this lower bound is at least 50% has been demonstrated by Newman and Waugh (1960), by Burton and Licklider (1955), and by Miller and Friedman (1957), in general agreement with Shannon. These later studies provide more stable estimates than Shannon gave. The evidence is clear that linguistic responses of human adults are tied closely to the redundancy of their language. Some of the most interesting and suggestive experiments in the use of sequential redundancy have been done by Chapanis (1954) and Miller and Friedman (1957, using random and regular deletions in texts. For example, in one of their "coherently mutilated" texts, Miller and Friedman showed that with an average abbreviation of 58% *S*s were able to reproduce 93% of the omitted letters.) A number of experiments (Aborn & Rubenstein, 1952; Miller, 1958; Pollack, 1953; Rubenstein & Aborn, 1954) establish that redundant sequences are easier to learn than random sequences. However, if learning scores are expressed as information per word, more information is gained in a unit of time for random sequences. Recent work on the acquisition of grammar by children (Braine, 1963) has established that some rather complex and surprisingly subtle syntactic constraints are to be found in the generative grammars induced from a speech corpus of a child only 30 mo. old. It is thus apparent that the structure of the material to be learned is an important variable.

The present study had the immediate aim of assessing the ability of children to deal with sequential redundancies and to gain some notion of the rate at which this ability grows. It was reasonable to suppose that redundant strings with their more coherent structure would be easier for children to learn than random strings. On the other hand, the more commerce a child has with language the more ought his coding strategies to generalize to redundant strings. Thus with redundant strings older children should learn faster than younger children. Random strings should be less amenable to recoding and as a consequence the ability of older children to learn random strings should be little if any greater than that of younger children.

---

[1] The authors express their appreciation for the outstanding cooperation of the Iowa City Board of Education and Charles L. Railsback, Principal of Longfellow Grammar School. These studies were supported in part by a grant from the United States Public Health Service (MH-06986-01).

The strings of symbols used in these experiments were identical with those produced by Miller's (1958) finite state generator. This generator is an algebraic system that selects a proper subset of strings from the set of all possible strings that can be formed from a set of symbols. Specifically, Miller used as symbols four randomly chosen consonants (G, N, S, and X). The set of all possible different strings of Length 2 that can be formed from these four symbols is just 4^2. For our purposes it is enough to note that the capacity of this redundant strings generator can be obtained by standard methods. Its capacity is .552 bits/letter. The system generating all possible strings using these four symbols has a capacity of $\log_2 4 = 2$ bits/letter. The redundancy in percent is thus

$$100\left(1-\frac{\text{capacity of Redundant generator}}{\text{capacity of Random generator}}\right)=72.4,$$

and the redundant strings are 2/.552 = 3.62 times as long as random strings that encode the same information.

METHOD

Subjects.—The *S*s were 192 children, 64 from each of the fourth, fifth, and sixth grades. A child was assigned at random to one of 16 conditions obtained by permutation of two ordered lists of strings of symbols from the four lists used, two of which were redundant (L) lists and two of which were random (R) lists. The same number of *S*s from each grade served in each of the 16 conditions. Each condition contained the same number of *S*s. An additional dimension orthogonal to both grade and condition was the list material, the lists being composed of either strings of letters or strings of digits. Half the *S*s learned the strings used by Miller (1958) while the other half learned the strings produced by replacing Miller's consonant letters S, X, N, and G by randomly selected digits 9, 8, 2, and 6, respectively.

Materials.—The four basic lists of strings were the two different random lists, R_1 and R_2, and two different redundant lists, L_1 and L_2. There were eight lists in all because for each of these lists one identical in structure was obtained by substituting digits for letters. Every list was made up of nine different strings: one four-symbol string, two five-symbol strings, two six-symbol strings, and four seven-symbol strings. The nine strings in a list were randomized to give a different ordering of the strings for each of the 10 successive presentations. The same random ordering was used in each list. This controlled the effects of variable string length in different positions. Thus for all lists the four-letter strings appear at the same points in the 90-string sequence, and similarly for the other string lengths. The strings were drawn in black ink on 4 × 6 in. cards and the eight lists were photographed, with two blank frames interposed between each of the 10 nine-frame trials.

An *S* made his responses on every other page of a booklet of lightweight 3 × 5 in. cards, following each showing of a nine-frame trial onto a piece of black construction paper by a Dunning Animatic strip film projector. The strings were seen as black on a light gray field surrounded by a border of black. The gray field was 236 mm. wide and 162 mm. high, the letters were 7 mm. high, and the black background was 454 mm. wide and 300 mm. high.

Learning trials.—Three *S*s, one from each grade, were run in each session. They were seated together at a table with their heads approximately 2.16 m. from the paper screen. A session consisted of 10 training trials on an initial list, 10 trials on a second (not always different) list, and a recall trial during which no strings were shown. Each trial began with the presentation of each of nine strings for a 6-sec. period with no interstring (interframe) interval, and was terminated by an unpaced recall period. The *S*s were instructed to write at each recall period as many of the strings they had just been shown as they could remember. They were encouraged to guess when in doubt, and were told that they could write the strings in any order they chose but that they should try to write the letters (numbers) in the correct order in each string that they wrote. The *S*s were instructed not to write anything while the strings were being presented. Each *S* turned 2 pages over after finishing writing and *E* began the next trial after all *S*s had finished. The *S*s were told nothing about the construction of the lists.

Following the training session *S*s were instructed either that a new list would be shown next following a momentary rest (two L_xL_y groups, two R_xR_y groups, four LR groups, and four RL groups, where $x \neq y$), or that after a momentary rest they would continue (Groups L_1L_1, $L_2L_2 = L_xL_x$; Groups R_1R_1, $R_2R_2 = R_xR_x$). A new film strip, or the same strip, was then threaded and the transfer session was run. Immediately after the tenth trial of the second session *S*s in the RR and LL conditions were instructed to write down, once again, as many strings as they could. All other *S*s were told to write down as many strings as could be remembered from the first list.

RESULTS

The mean number of strings correctly recalled as a function of trial number is shown in Fig. 1 (letters) and Fig. 2 (digits). The left hand graphs of each figure show, for Grades 4, 5, and 6, the data for those *S*s who learned a redundant (L) list first or a

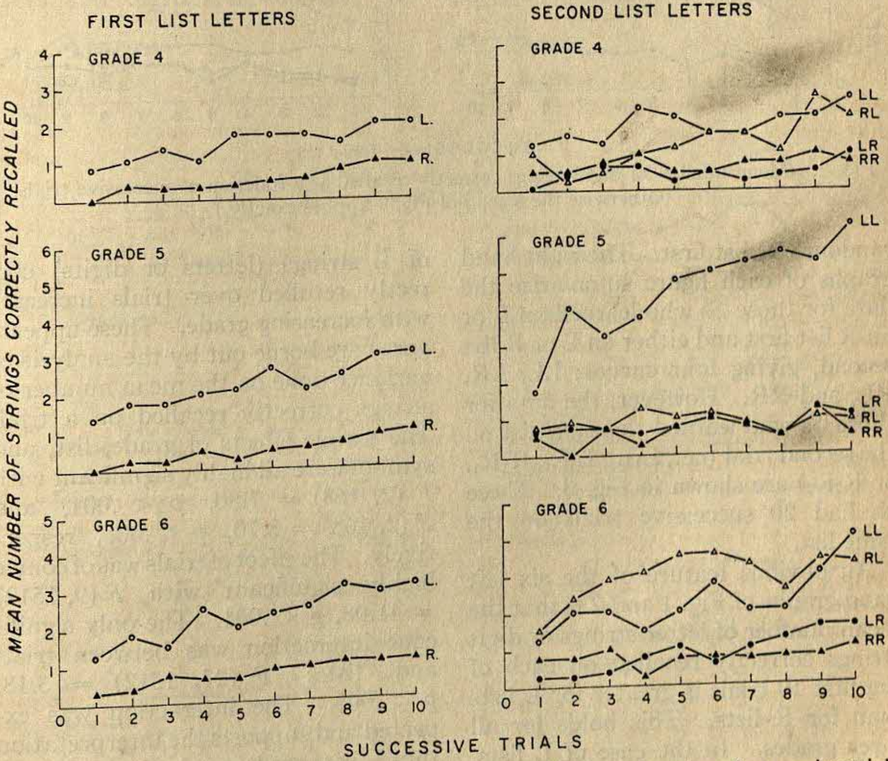


FIG. 1. Mean number of letter strings correctly recalled as a function of successive trials. (A redundant list is indicated by L, a random list by R. First list graphs L show results for the 16 *S*s who studied RR or RL. The second list graphs are based on 8 *S*s for each of the RL or LR order and on 4 *S*s each for the L_1L_2, L_2L_1, R_1R_2, and R_2R_1 orders. The results for *S*s—4 in each case—whose order was L_1L_1, L_2L_2, R_1R_1, and R_2R_2 are shown in Fig. 3.)

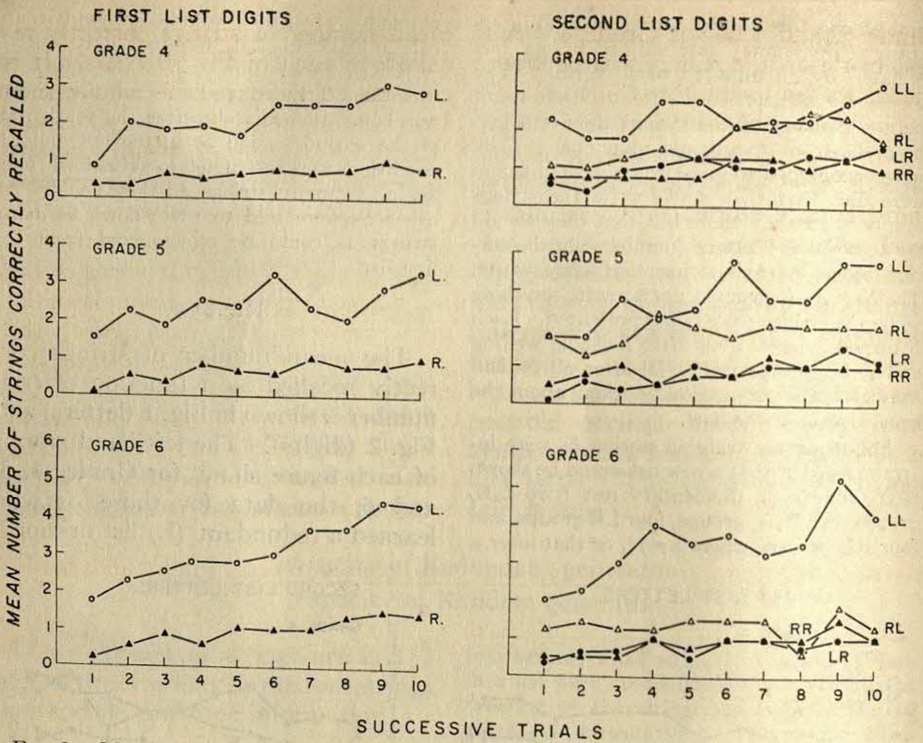


FIG. 2. Mean number of digit strings correctly recalled as a function of successive trials. (Otherwise the legend of Fig. 1 applies exactly.)

random (R) list first. The right hand graphs of each figure summarize the data for those *S*s who learned an L or an R list first and either an L or R list second, giving four curves: LL, LR, RL, and RR. However, the data for those *S*s who learned the *same* list on the second trial (i.e., L_1L_1, L_2L_2, R_1R_1, or R_2R_2) are shown in Fig. 3. These *S*s had 20 successive trials on the same list.

An obvious feature of the six left hand graphs of Fig. 1 and 2 is that the mean number of letter strings or digit strings correctly recalled on each of the first 10 trials is greater for L lists than for R lists. This holds for all three grades. In the case of L lists, more digit strings than letter strings are correctly recalled on a given trial number, and this difference increases with trial number. The mean number of L strings (letters or digits) correctly recalled over trials increases with increasing grade. These appearances are borne out by the analysis of variance done on the mean number of strings correctly recalled on a trial. The simple effects of grade, list, and symbol were all highly significant with $F\ (2, 168) = 7.80$, $p < .001$, and $F\ (2, 168) = 8.70$, $p < .005$, respectively. The effect of trials was of course highly significant with $F\ (9, 1512) = 41.98$, $p < .001$. The only significant interaction was between trials and lists: $F\ (27, 1512) = 3.18$, $p < .005$. The interaction was expected and supports the interpretation that L lists are learned at a faster rate than R lists. The mean number of R strings (data for letters and digits pooled) correctly recalled over the 10 trials was .78 for Grade 4, .87 for

Grade 5, and 1.00 for Grade 6. A *t* test made on the differences between means was not significant.

The grade school *S*s in this study hardly performed at the level of Miller's (1958) *S*s (Harvard faculty and graduate students) whose mean recall scores were 3.5 R strings and 9 L strings, or at the level of 54 UCLA Freshmen, run in a preliminary study using the letter strings, whose mean correct recall scores were 2.5 R strings and 7.8 L strings. They do, however, show the same marked superiority in recall of redundant strings over random strings.

The three right hand graphs of Fig. 1 and 2 show for second-list letters and digits, respectively, the mean number of strings correctly recalled on each of the 10 trials. It is clear that the groups of *S*s who learned L strings on the second 10-trial session (LL and RL) generally recall correctly more strings than RR and LR groups. The LL groups were superior to RL groups. The learning of a second R list was equally difficult whether the list learned first was a redundant (L) or a random (R) list. Also, the performance of LL groups clearly improves as grade increases.

Figure 3 plots the mean number of strings of letters and digits recalled on 20 successive trials for those *S*s who received the same list on both sessions of 10 trials each. In general, LL lists are learned more rapidly and better

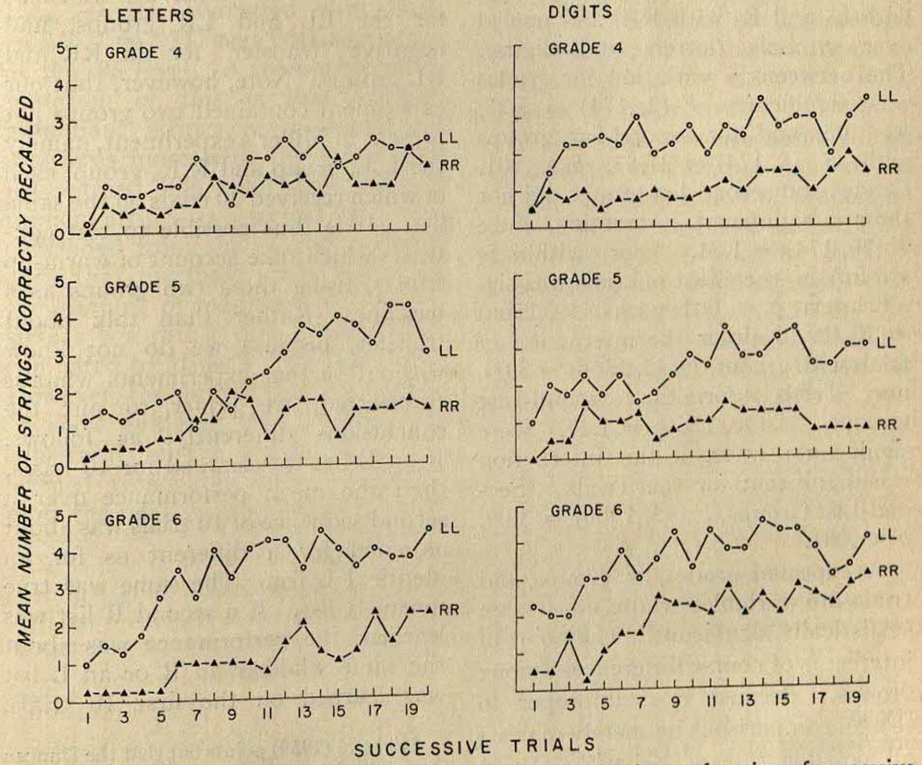


FIG. 3. Mean number of strings of letters or digits recalled as a function of successive trials for those *S*s whose second list was identical to the first, that is, for the conditions L_1L_1, L_2L_2, R_1R_1, and R_2R_2. (There were four *S*s in each of these groups in each grade and symbol condition. Thus a given *S* had 20 recall trials on the same set of strings.)

than the RR lists. There is little improvement in either type of list after the tenth trial. Performance with digits is slightly better than for letters, and the mean number of strings correctly recalled over the entire 20 trials increases with grade. The data of these groups are a baseline against which to assess the performance of the four groups whose first and second lists were different.

An analysis of variance was done on the recall scores for the 10 trials with the second lists. The basic data were the mean number of strings correctly recalled on a trial (1–10) by each grade (4, 5, and 6) for each second-list group (L_xL_x, L_xL_y, LR, R_xR_x, R_xR_y, and RL) pooled within different lists for the RL and LR groups (i.e., L_1 with L_2 and R_1 with R_2), and pooled over symbols (letters and digits). The between-*S*s variation for grades was significant, $F\ (2, 174) = 3.27$, $p < .05$, and for second-list groups with $F\ (5, 174) = 21.57$, $p < .001$. Grades and second-list groups did not show a significant interaction, since $F\ (10, 174) = 1.21$. For within-*S*s variations, the effect of trials was significant at $p < .001$, with $F\ (9, 1566) = 20.40$. Neither the interaction of trials with grades, $F\ (45, 1566) = 3.09$, nor Trials × Grades × Second-List Groups, $F\ (90, 1566) = 1.27$, were significant; whereas the interaction was significant for the Trials × Second-List Groups, $F\ (45, 1566) = 3.09$, $p < .001$.

As expected, grades, list groups, and trials are variables whose effects are statistically significant. The point of interest is of course differences among groups. Because it is improper to test all comparisons by merely making all possible *t* tests, Scheffé's (1959) *S* method was used for judging all contrasts that may be of interest, "including any suggested by the way the data fall out [p. 66]."[2] On the basis of a 95% overall confidence coefficient, it was concluded that any LL group has a higher mean recall score on the second list than any LR, RL, or RR group; the RL group's recall score is higher than that of the LR group; and no other differences are statistically significant.

Miller (1958) found positive transfer for the LL, LR, and RR groups and negative transfer for the RL group. His criterion for transfer appeared to be a comparison between the mean number of strings correctly recalled on the tenth trial of the second list with that score on the tenth trial of the first list. Using this criterion for the present experiment we should find positive "transfer" for the LL and LR groups, and negative "transfer" for the RR and RL groups. Note, however, that our experiment contained two groups not present in Miller's experiment, namely the L_xL_x group and R_xR_x group, each of which received 20 trials on the same list. It is thus possible to test contrasts which take account of warm-up effects, using these two groups as a baseline. Rather than talk about transfer, because we do not know without further experiments what is transferred, we prefer to put the conclusions differently, as follows. If an L list was learned for 10 trials, then the mean performance over a second sequence of 10 trials was about as good for a different as for an identical L list. The same was true for an R list. If a second R list was learned, the performance was about the same whether an R or an L list was learned on the first 10 trials.

---

[2] Scheffé (1959) points out that the Duncan range test for making all possible comparisons among, say, a set of means, rests on poorly understood assumptions. He discusses other alternatives including his own very general *S* method and the *T* method, due to Tukey.

However, a significant decrement was found in learning an L list second if an R list was learned first. This finding is in accord with Miller's (1958) conclusion that,

> The strategy that *S*s adopted for memorizing redundant strings apparently did not handicap them when they later encountered random strings, but the reverse is not true: The strategy adopted for the random strings was not efficient when later applied to redundant strings [p. 490].

What is the effect of learning of the second list on recall of strings from the first list? The analysis of variance of the recall data of Table 1 showed that the only significant effect was due to groups, with $F\,(3, 120) = 22.59$, $p < .001$. The t tests done on L_xL_y compared with LR ($t = 2.35$) and L_xL_y + RL compared with R_xR_y + LR ($t = 2.22$) were both significant at the .05 level. A reasonable interpretation of this is that learning redundant strings as a second list interferes more with the recall of a list learned first than does learning random strings as a second list. It is concluded from the overall performance that for these grade school *S*s the learning of redundant or random strings is a rather slow process when compared to Miller's (1958) adult *S*s. However, the printed responses of the *S*s show evidence that at least one constraint in the strings has been learned reasonably well. Figure 4 plots the relative proportions of recall attempts ending in "G" or "6" as a function of trials, for the four groups L_xL_y, R_xR_y, RL, and LR. The crucial changes come during the second 10 trials. Against the symptotic levels of L_xL_y and R_xR_y, the LR group moves from about .94 to .35 (on the tenth recall trial) and the RL group from about .20 to about .80. Since the proportions of "G" (or "6") endings are 1.00 for the R strings and .33 for the L strings, it seems clear from the several curves of Fig. 4 that *S*s very quickly come to respond in direct proportion to one of the sequential constraints. The RL and LR curves show likewise that the effect of constraints in the last symbols of the strings learned during the first 10 trials is not dispelled in the course of 10 trials on the second strings, although the curves appear to be approaching the L_xL_y and R_xR_y asymptotes, respectively.

These conclusions are borne out by the analysis of variance done on an arc-sine transformation of proportion of "G" and "6" response made during the 10 second-list trials by each *S*. A simple randomized design was used over the four groups R_xR_y, RL, LR, and L_xL_y, with scores pooled over all other dimensions. The effect of groups was significant at $p < .001$, $F\,(3, 143) = 109.49$. The same transformation and analysis of variance were done on the proportion of re-

TABLE 1

MEAN NUMBER OF STRINGS RECALLED FROM FIRST LIST AFTER STUDYING THE SECOND LIST

Grade	Letters			Digits			Overall *M*
	4	5	6	4	5	6	
L_xL_y	0.25	1.25	0.50	0.75	2.25	1.75	1.13 (5.0)[a]
LR	1.37	2.25	2.00	1.38	1.88	1.63	1.73 (5.0)[a]
RL	0.13	0.13	0.38	0.13	0.25	0.00	0.13 (1.8)[a]
R_xR_y	0.50	0.00	0.75	0.00	0.25	0.50	0.33 (0.5)[a]

[a] Data from Miller (1958).

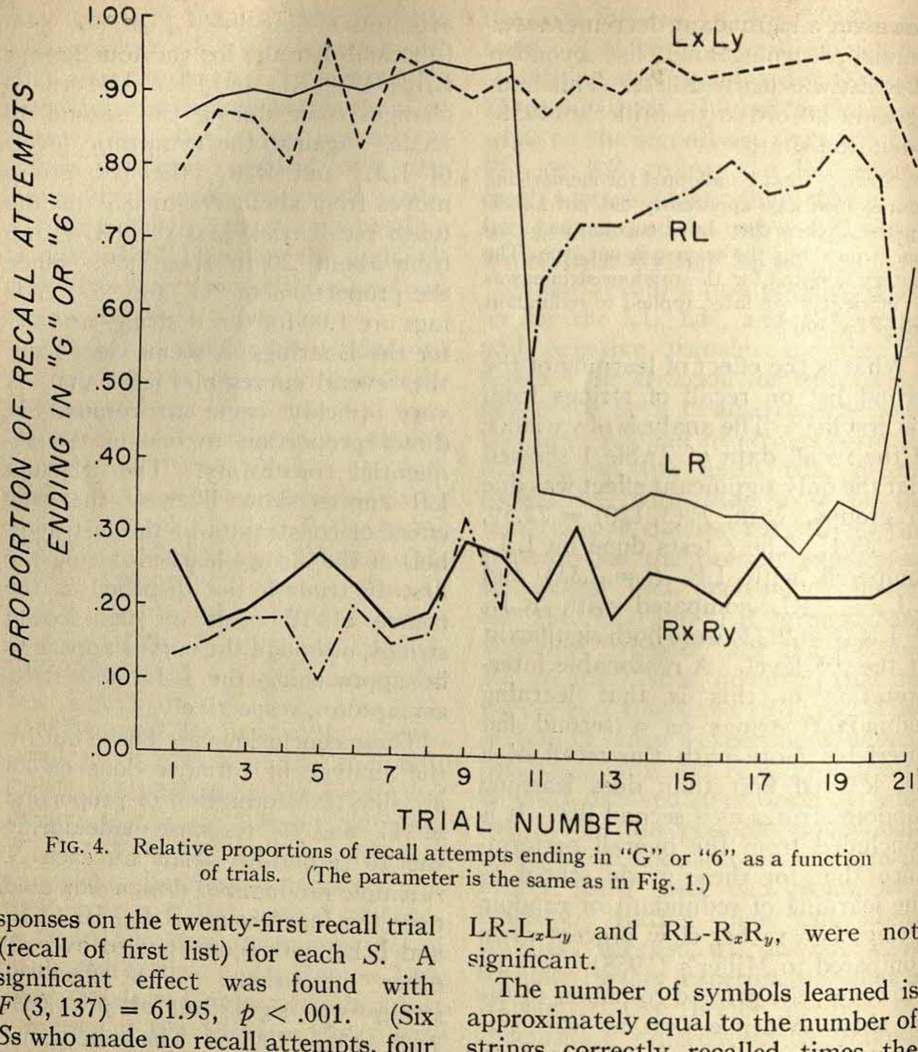


FIG. 4. Relative proportions of recall attempts ending in "G" or "6" as a function of trials. (The parameter is the same as in Fig. 1.)

sponses on the twenty-first recall trial (recall of first list) for each *S*. A significant effect was found with $F(3, 137) = 61.95$, $p < .001$. (Six *S*s who made no recall attempts, four in the RL and two in the LR groups, and their data were omitted from the analysis.) Since the *F*s were both significant, the differences of interest during second-list trials were between R_xR_y and LR, and between RL and L_xL_y. The *t* tests made on these two differences ($df = 140$) were 1.49 ($p < .10$) and 5.07 ($p < .0005$), respectively. Thus RL is significantly lower than L_xL_y, while R_xR_y and LR do not differ. This provides additional evidence for interference in the case of the RL group. The differences of interest for the recall trials, LR-L_xL_y and RL-R_xR_y, were not significant.

The number of symbols learned is approximately equal to the number of strings correctly recalled times the average length of a string. The first half of Table 2 gives the mean number of symbols learned in the first 10 trials by letters and digits, and by grade. Translating this into mean number of bits by multiplying the mean number of L and R symbols learned by .552 and 2 bits, respectively, the second part of Table 2 is obtained. Dividing each of the mean bits quantities by 10 will give the number of bits/trial. Except for one case—digits, Grade 4—with redundant strings, *S*s learn more symbols but less information than with random strings. The average information learned per trial is greater for digits than for letters, but the difference between R and L strings is greater for letters than for numbers. Miller (1958), too, found this result for letter strings. Aborn and Rubenstein's (1952) conclusion is thus reconfirmed for grade school *S*s with highly redundant strings learn more symbols but less information than with random strings.

DISCUSSION

It is essential to consider here Garner's (1962) critique of the original Miller (1958) study, published after the completion of our experiments. Garner holds that Miller's experiment (and therefore ours) is an experiment on the form of redundancy, not on the amount of it and, in fact, that the *total amount* of constraint in the constrained and random lists is identical. It is only the *form* of the constraint that is different. This is because the constrained and random lists of words each contained the same number and lengths of words, and with four alternative letters the total number of words in the parent population of words is the same. Garner (1962) states

> The same total amount of constraint can occur in many different forms, and *the form of the structure is not determined by how many patterns are used out of the total possible, but rather by which particular patterns are used* [italics his] [p. 157].

In general, Total Constraint = Internal Constraint+External Constraint. Garner does admit that there probably are not perfect psychological performance counterparts of internal and external constraint, although it is clear that they are to be identified with free recall learning and perceptual discrimination, respectively. We can, then, interpret the present experiments as bearing on the form of redundancy—the L lists are clearly more constrained than the R lists. On this view our findings are unaltered but the conclusion must be modified to apply to the form of the redundancy rather than to its amount.

TABLE 2

MEAN NUMBER OF SYMBOLS AND BITS LEARNED IN THE FIRST 10 TRIALS

List	Grade 4	Grade 5	Grade 6
	Mean Number of Symbols		
Letters			
R	6.0	6.0	6.7
L	12.8	18.4	18.8
Digits			
R	4.0	6.0	8.6
L	17.3	19.9	26.2
	Mean Number of Bits		
Letters			
R	12.0	12.0	13.4 (41.0)[a]
L	7.1	10.1	10.4 (29.8)[a]
Digits			
R	8.0	12.0	17.3
L	9.5	11.0	14.5

[a] Data from Miller (1958).

Recently Whitman and Garner (1962) studied explicitly the role of the form of internal structure in free recall learning, comparing subsets of visual figures all drawn from a larger set, but having different forms of internal structure. They concluded that (*a*) the ease of free recall learning depends on the characteristics of the entire subset to be learned rather than on those of the individual stimuli and that (*b*) internal structure which exists in the form of simple contingencies between variables is better for free recall than are more complex forms of structure. As they point out, it is not easy to state the amount of simple contingency in our lists because not all strings in a list are of the same length. However, the nature of the differences that exist between the R and L lists appears to be similar to the contingency differences used in that study. The following attempt was made to estimate the amount of simple contingency in the R and L lists.

The relative sequential constraint (Garner, 1962) in sequences of pairs of letters ($x_1x_2, x_2x_3 \ldots x_{n-1}x_n$) was computed for the L_1 and R_1 lists, under two different assumptions: (*a*) That the strings formed a continuous sequence of symbols with a "space" symbol between strings, or (*b*) That a list of strings formed a single continuous sequence of symbols with no "space" symbol. Assumption *a* led to a value of relative sequential constraint, $U(1:2)/U(1)$, of $1.16/2.19 = .53$ for the L_1 list and $(.36/2.29) = .16$ for the R_1 list, where $U(1:2)$ is the (simple) contingent uncertainty in the prediction of the second of a sequence of two letters, and $U(1)$ is the uncertainty of the first position. $U(1:2)$ is a measure of the amount of uncertainty reduction due to the contingency or correlation between Positions 1 and 2 in a sequence of two letters. The corresponding values for Assumption *b* were $(.60/1.86) = .32$ for L_1 and $(.203/1.997) = .10$ for R_1. It is seen that the relative sequential constraint in the random list is not negligible, being about one third of that in the redundant list under either assumption.

The ratio of the amount (in bits) of simple contingencies in two subsets of the Whitman and Garner (1962) study was exactly three (4.74/1.58). Their *S*s' free recall performances on the high- and low-contingency subsets paralleled that of our *S*s on the L and R lists. It is a reasonable conclusion, then, that one probable basis for the faster learning by our *S*s of the redundant lists is their greater degree of simple contingent constraint relative to that of the random lists.

The general outcome of these experiments is to confirm Miller's (1958) findings with grade school children as *S*s, and extend them to random and redundant strings of digits. The similar results for letters and digits indicate that structure, as defined by the total set of patterns and the sequential constraints between the strings and within the symbols of the strings, is an important determiner of learning in free recall. With increasing age the ability to freely recall lists of strings of symbols with high internal structure increases considerably between age 12 and adulthood. There is a very slight increase in an adult's free recall ability over that of children for the less constrained random lists. Now, we found that the higher was *S*'s grade, the faster either type of string was learned. However, it had been expected that the *differences* between rates of learning L and R lists would increase with grade. This effect would have shown itself as a significant interaction of grades and lists over trials. Either this effect does not occur, or is masked by taking place in the first trial or two, or a spread of more than three school grades is required to measure it.

REFERENCES

ABORN, M., & RUBENSTEIN, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44**, 260–266.

BRAINE, M. D. S. On learning the grammatical order of words. *Psychol. Rev.*, 1963, **70**, 323–348.

BURTON, N. G., & LICKLIDER, J. C. R. Long-range constraints in the statistical structure of printed English. *Amer. J. Psychol.*, 1955, **68**, 650–653.

CHAPANIS, A. The reconstruction of abbreviated printed messages. *J. exp. Psychol.*, 1954, **48**, 496–510.

GARNER, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

MILLER, G. A. Free recall of redundant strings of letters. *J. exp. Psychol.*, 1958, **56**, 485–491.

MILLER, G. A., & FRIEDMAN, E. A. The reconstruction of mutilated English texts. *Inform. Control*, 1957, **1**, 38–55.

NEWMAN, E. B., & WAUGH, N. C. The redundancy of texts in three languages. *Inform. Control*, 1960, **3**, 141–153.

POLLACK, I. Assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66**, 421–435.

RUBENSTEIN, H., & ABORN, M. Immediate recall as a function of degree of organization and length of study period. *J. exp. Psychol.*, 1954, **48**, 146–152.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

SHANNON, C. E. A mathematical theory of communication. *Bell Sys. tech. J.*, 1948, **27**, 379–423, 623–656.

WHITMAN, J. R., & GARNER, W. R. Free recall learning of visual figures as a function of form of internal structure. *J. exp. Psychol.*, 1962, **64**, 558–564.

(Received November 15, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 409–418

REVERSALS PRIOR TO SOLUTION IN CONCEPT IDENTIFICATION[1]

GORDON BOWER AND THOMAS TRABASSO[2]

Stanford University

These studies investigated the effects of reversal and nonreversal shifts before solution upon performance in a later concept identification task. In 2 experiments, a reversal or a nonreversal shift after an error on a critical trial had no interfering effect upon subsequent learning. The reversal and nonreversal groups made about the same number of errors and required as many trials to learn as did controls who were not shifted. In a 3rd experiment, 1 group of Ss received reversals on every alternate error, but still made the same number of informed errors as did controls who learned with no shifts. These results support the hypothesis that learning is insightful or an all-or-nothing event in simple concept identification.

In the typical two-category concept identification experiment, *S* is shown a series of complex patterns which vary in several, binary attributes. As each pattern is presented, *S* attempts to anticipate the correct classification; following his response, he is informed of the correct response. The patterns are divided into two mutually exclusive classes, R_1 and R_2. If, say, color (red or blue) is the relevant attribute, then red objects might be assigned to Response Class R_1 and blue objects to Class R_2. We will refer to this rule as a particular S-R assignment.

In recent studies (Bower & Trabasso, 1963a; Trabasso, 1963) of this situation with college students, *S*s appeared to learn suddenly. Backward learning curves were horizontal at the chance level of 50% correct classifications over all trials until *S*'s last error before solving. The performance of an *S* might be characterized by saying that on any given trial he is either in the presolution state or in the solution state, with corresponding probabilities of .50 or 1.00 of correctly classifying the stimuli. According to this two-state description of the performance, learning would be identified as a discrete, one-trial transition from the initial, presolution state into the terminal, solution state.

The theories of cue-selection learning proposed by Restle (1962) and Bower and Trabasso (1963a) imply this two-state description of individual performance. These theories assume that *S* is selectively attending to or sampling cues from the stimulus display and that he is testing hypotheses regarding the relevance of these cues to the correct solution. If *S*'s response is correct, it is supposed that he continues to use the same hypothesis; if his response is incorrect, he resamples at random from the set of possible hypotheses. Assume further that the proportion of correct hypotheses is c whereas the remaining proportion $1 - c$ is irrelevant hypotheses which lead to correct and incorrect responses half the time. By these assumptions, the probability that *S* solves the problem after any given error is a fixed constant, *c*. This elementary theory has been used successfully in predicting quantitative details of several sets of data (Bower & Trabasso, 1963a).

---

[1] This research was supported by Grant M-3849 to the first author from the National Institute of Mental Health. The same agency provided the second author with a postdoctoral fellowship, MPD-18070, during which time the work was done. Thanks are due to Milton Kielsmeier who assisted us in obtaining *S*s for Exp. III.

[2] Now at University of California, Los Angeles.

The present studies investigate whether *S* acquires partial knowledge about the solution to the problem. The all-or-nothing theory supposes that he does not. Specifically, it says that when *S* makes an error, he has not yet learned anything of relevance regarding the correct concept. Three experiments were performed to provide tests of this assumption; the first two are described now.

EXPERIMENTS I AND II

Experiments I and II are identical in design; Exp. II was a replication of Exp. I with an easier problem and different stimulus materials. The design resembles that used in several animal experiments conducted on the continuity-noncontinuity issue in discrimination learning theory (e.g., Krechevsky, 1938; McCulloch & Pratt, 1934). Control *S*s in Group C learned a problem with the same S-R assignments throughout (Cue A-R_1, Cue B-R_2). Two other groups worked on different S-R assignments initially and then were transferred to the assignments of the control group. This transfer occurred immediately after *S* made an error following a critical trial of the initial series. Group R, a reversal group, was trained initially with the opposite assignments, A-R_2 and B-R_1. Group NR, a nonreversal-shift group, was trained initially with Cues A and B present but irrelevant while another set of cues was relevant (C-R_1, D-R_2).

The question of interest is whether the initial wrong-way training retards performance of *S*s in Groups R and NR who are shifted to the final, transfer problem before solving their initial problem. If *S*s partially learn responses to the initially relevant cues before the shift, then such partial learning should induce negative transfer on the final problem. However, if *S*'s error initiating the shift indicates that nothing of importance has yet been learned, then the performance on the final problem should be the same for the three groups, independent of the initial S-R assignments.

Method

Experimental design.—A schematic outline of the design is presented in Table 1. Only two of the several stimulus attributes are represented in the left columns of Table 1. The rows give the combinations of stimulus values in the patterns and the correct responses to each pattern are listed under each condition. The Control and Reversal groups had Cues A and B relevant but they had opposite response assignments during initial training (10 trials in Exp. I and 5 trials in

TABLE 1

DESIGN FOR EXP. I AND II

Patterns		Response Assignments			
		Initial Trials			Final Problem
Dimension 1	Dimension 2	Control	Reversal	Nonreversal	
A	C	R_1	R_2	R_1	R_1
A	D	R_1	R_2	R_2	R_1
B	C	R_2	R_1	R_1	R_2
B	D	R_2	R_1	R_2	R_2

Exp. II). The Nonreversal group had one of the other dimensions (Cues C and D) relevant during initial training.

The *S*s who made an error on Trial 10 in Exp. I or Trial 5 in Exp. II or soon thereafter were immediately shifted to the final problem listed in the right hand column of Table 1. We wished to compare on this final problem only those *S*s who had not yet learned their initial problem by Trial 10 (or 5 in Exp. II). Consequently, if an *S* in any group began a criterion run of 16 consecutive correct responses on or before the critical trial (10 or 5), he was not shifted but was, as a result, excluded from the critical comparison between those *S*s who did get put onto the final problem. According to the theory, these latter *S*s were equalized at the start of the final problem since each *S* made an error before the shift was effected.

Procedure.—The same instructions were read to all *S*s. The *S* was to classify a set of patterns into two classes. In Exp. I, the classificatory responses were MIB and CEJ; in Exp. II, the numerals 1 and 2. The *S* was told that the patterns could be classified by a simple principle.

Patterns were presented one at a time on a card holder. The *S* paced his verbal responses and *E* then stated the correct classification. The *S* was allowed 4 sec. to view the pattern after reinforcement. A different order was presented each *S* by shuffling the cards before the session. Cards were reshuffled at the end of every 64 trials if *S* had not yet reached the learning criterion of 16 successive correct responses.

Stimulus materials.—For Exp. I, patterns were constructed by sampling a single letter from each of four pairs of letters, (V or W), (F or G), (X or Y), (Q or R). Thus, VFYQ was a pattern, WVXR was not. The four letters were printed in a diamond shape on a 3 × 5 in. card. The letters appeared fixed in the order given above, but their locations at the four diamond corners rotated randomly from trial to trial. Location was an irrelevant cue. For Groups C and R, the letter pair (V, W) was relevant; the classification depended on which one of the letters was present on the card. One of the other letter pairs was selected randomly to be initially relevant for each *S* in Group NR, whereas (V, W) was irrelevant. The final problem was with (V, W) as the relevant cues with response assignments V-MIB and W-CEJ.

For Exp. II, the stimuli were geometric figures drawn in crayon pencil from templates on white 3 × 5 in. file cards. There were six binary dimensions: color (red or blue); size (large or small); shape (square or hexagon); number (three or four figures); position (figures arranged along right or left diagonal); and colored area within each figure (upper-right and lower-left or upper-left and lower-right quadrants). There was one relevant dimension and five irrelevant dimensions for each group. Color was relevant for Groups C and R. One of the other five dimensions was randomly selected and made relevant during initial training for each *S* in Group NR.

Subjects.—For Exp. I, the *S*s were 65 students in the introductory psychology course at Stanford University. Eleven *S*s began a criterion run on or before Trial 10; there were 4, 3, and 4 *S*s in Groups C, R, and NR, respectively. These *S*s do not enter into the comparison on the final problem since they were not transferred. Setting aside these *S*s, 18 *S*s (13 males and 5 females) remained in each group.

For Exp. II, the *S*s were 46 students in the introductory psychology course at Stanford University. Since the problem was easier, a larger proportion of *S*s was expected to solve within a few trials. Hence, fewer initial training trials (five) were used so that the majority of *S*s would not have to be set aside. Sixteen *S*s, 5 in Group C, 4 in Group R, and 7 in Group NR, began a criterion run on or before Trial 5. These *S*s were excluded from comparisons on the final problem. There remained 10 *S*s (6 males and 4 females) in each group for comparison on the final problem.

Results

In Exp. I, one *S* in Group C and one in Group NR failed to reach criterion within 140 trials on the final problem; all other *S*s solved within 140 trials. In Exp. II, all *S*s solved the final problem. Comparisons among groups on final-problem performance refer to trials following the error trial that initiated the shift to the final problem for a given *S*. Average errors and trial of last error are shown in Table 2 for the three conditions in both experiments.

The group differences on mean errors and mean trial of last error on the final problem were negligible in both experiments. The learning-parameter estimates (reciprocal of mean

TABLE 2

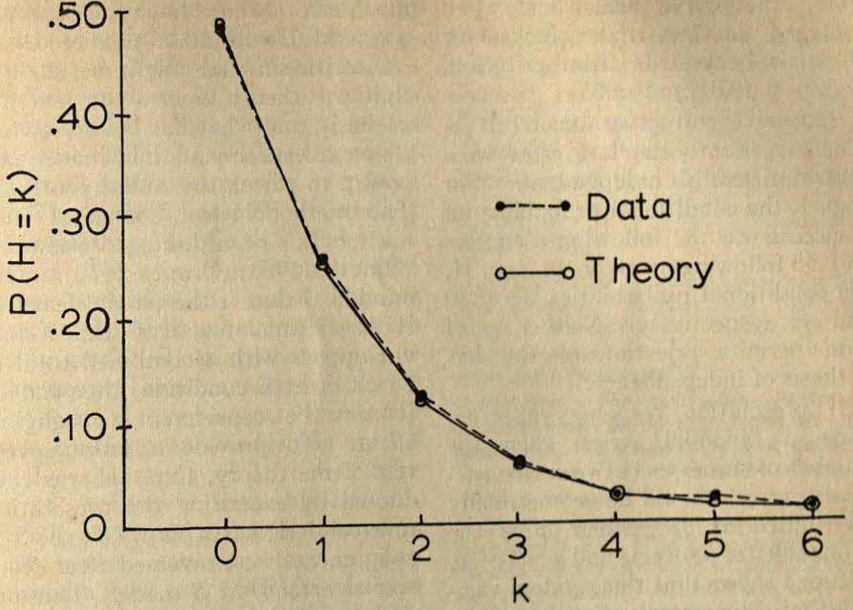
MEAN ERRORS AND TRIAL OF LAST ERROR, *SD*s, AND c ESTIMATES FOR THE FINAL PROBLEM

Group	N	c	Mean Errors	SD	Mean Trial of Last Error	SD
Exp. I						
Control	18	.052	19.11	19.01	38.33	32.50
Reversal	18	.052	19.11	16.42	39.56	32.27
Nonreversal	18	.055	18.28	19.28	36.94	38.23
Exp. II						
Control	10	.078	12.90	8.42	28.60	20.82
Reversal	10	.067	14.90	9.77	29.00	19.71
Nonreversal	10	.071	14.00	14.15	26.90	26.45

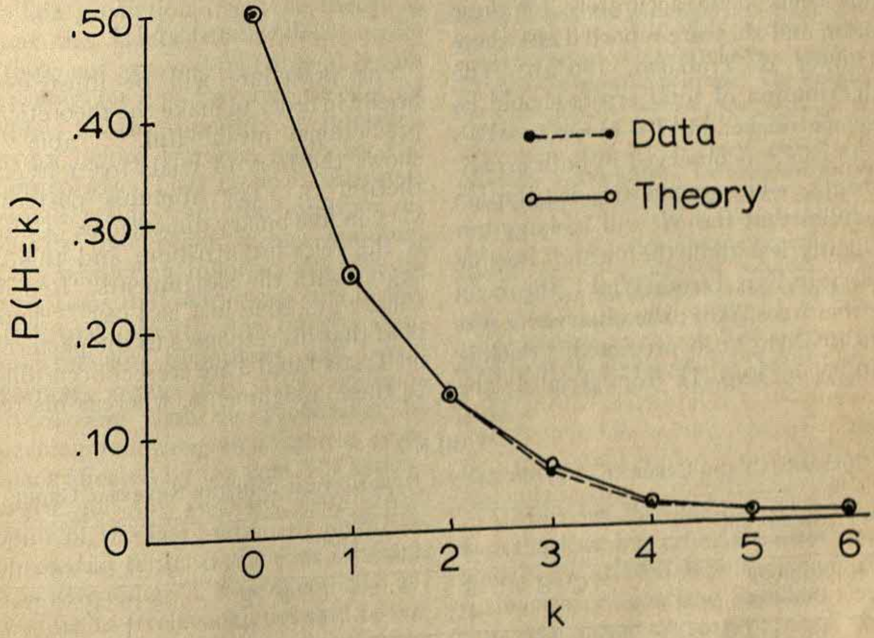
errors) are shown in Table 2; a likelihood ratio test for equality of c's was nonsignificant in both experiments. Further, a likelihood ratio test that each *S*'s learning parameter, c_i, was equal to a common c was tested for all 65 *S*s in Exp. I and for all 45 *S*s in Exp. II. In each case, the null hypothesis could not be rejected—for Exp. I, χ^2 (64) = 53.3, $p > .05$; for Exp. II, χ^2 (45) = 42.4, $p > .05$ (Bower & Trabasso, 1963b). Thus, the data were consistent with the hypothesis of a common c for *S*s in each experiment; the differences among *S*s' error scores could be attributed to the variability inherent in the theoretical process.

The lack of group differences indicates that performance on the final problem was unrelated to the response assignments reinforced during the initial series. Correspondingly, there was no evidence for partial learning of the relevant cues or partial elimination of irrelevant cues (cf. Group NR). Effectively, we may rely upon a single error by *S* to indicate that he is "naive" about the correct solution. An error in this situation has the properties of an uncertain recurrent event (Restle, 1962); when *S* commits an error, we may, so to speak, reset him back to the starting point from which he began working on the problem. It should be noted that the null effects of reversal and nonreversal shifts before solution differ from the effects of such shifts after initial solution has occurred (Kendler & Kendler, 1962). What differs in the two cases is that after solution, *S* has a strong bias to attend to the formerly relevant cue, whereas before solution he is sampling cues at random to test (Kendler, Glucksberg, & Keston, 1961).

Presolution analyses.—The data prior to the last error of each *S* were analyzed according to the expectations of the all-or-none theory. In theory, the presolution responses of *S* may be represented as a stationary and independent binomial process. To test for a constant probability of success prior to the last error, backward learning curves (Hayes, 1953) were constructed. These were stationary near .50 in each experiment. Pooling the two experiments, the probabilities of a correct classification on Trials −1, −2 ··· −8 backwards from the last error were .50, .50, .52, .53, .54, .49, .51, and .50. Further-



1a. Data of Exp. I.



1b. Data of Exp. II.

FIG. 1. Distribution of number of successes intervening between two adjacent errors.

more, the curve was flat when analyzed in two-trial blocks for 40 trials backwards from criterion, $\chi^2\ (19) = 19.31, p > .05$.

Successive correct or incorrect responses prior to the last error were also statistically independent. For Exp. I, the conditional probability of a success was .52 following a success and .53 following an error; in Exp. II, the conditional probabilities were .50 and .52, respectively. Neither set of data permits rejection of the hypothesis of independence.

If presolution responses approximate a binomial series, then the number of successes between two successive errors should be geometrically distributed as qp^n, where p is the probability of a success and $q = 1 - p$. Figure 1 shows that this random variable has a geometric distribution in the data of the two experiments.

A number of numerical predictions has been made accurately for these data, and they are reported elsewhere (Bower & Trabasso, 1963a). The distribution of total errors should be geometric, i.e., $\Pr\{T = k\} = c(1 - c)^{k-1}$, and this was observed in both experiments. The geometric distribution implies that the *SD* will be large but slightly less than the mean. Pooling all *S*s in Exp. I from Trial 1, the mean errors were 20.85; the observed σ was 18.49 with 20.30 predicted. Pooling all *S*s in Exp. II from Trial 1, the mean errors were 11.45; the observed σ was 11.02 with 10.96 predicted.

A criticism that might be made is that the theory asserts the null hypothesis, and what has been shown is that our experiments had inadequate power to reject the null hypothesis. The methodological status of such matters has been discussed elsewhere (Binder, 1963; Grant, 1962). Our opinion is that if the partial learning is of such small magnitude that it does not appear with a combined total of 28 *S*s in each condition, then indeed it may be considered a negligible effect. To provide a more severe test of the theory, Exp. III was conducted by extending the presolution reversal design. In Exp. III, the S-R assignments were reversed after every second error that *S* made. Thus, as *S* proceeded along through his series of trials, the S-R assignments were switching repeatedly back and forth.

EXPERIMENT III

The procedure will be illustrated briefly in order to make the theoretical predictions meaningful. Table 3 shows the first 14 trials for a hypothetical *S*. The stimulus patterns vary in five binary dimensions. Color is the relevant attribute and this *S* begins with the assignments "Red in Class VEK, Blue in Class CEJ." Suppose that his responses to the patterns on Trials 1 and 3 are correct according to these assignments, whereas his re-

TABLE 3

CORRECT (C) AND ERROR (E) RESPONSES OF A HYPOTHETICAL *S* IN THE REVERSAL GROUP

Alternating S-R Assignments	Trials 1	2	3	4	5	6	7	8	9	10	11	12	13	14
Red-VEK Blue-CEJ	C	E	C	E ↓				C ↑	E	E ↓			C ↑	E
Blue-VEK Red-CEJ				C	C	E	C	E		C	E	C	E	

sponses to the patterns on Trials 2 and 4 are wrong. The second error, occurring on Trial 4, initiates an immediate reversal of the S-R assignments, and *S* is told "Correct" for his response on Trial 4. According to the reversed assignments, the responses on Trials 4, 5, and 7 are correct whereas those on Trials 6 and 8 are errors. The second error of this subseries, occurring on Trial 8, initiates another immediate reversal back to the original S-R assignments, and the response on Trial 8 is called "Correct." The series of reversals on every second error continues in this fashion until *S* produces a string of 10 consecutive correct responses since his last reversal. A second group of control *S*s was never reversed; they simply learned a fixed set of S-R assignments by the conventional training procedure in which they were informed of every error.

The prediction from the theory is that the number of *informed* errors (those not arrowed in Table 3) before learning for the Reversal *S*s will be equal to the number of informed errors before learning made by the Control *S*s. That is, no interference should result from the multiple reversals that occur during training. This prediction follows from the assumptions that learning occurs in one trial, that opportunities for giving up an irrelevant hypothesis, and hence learning, occur only after *informed* errors, and that the probability of learning following an informed error is not affected by the past S-R assignments for the relevant cues. The point of the last statement might be phrased in terms of the cues to which *S* is attending: if *S* is selectively attending to irrelevant cues and is not "noticing" the color, then his behavior when he starts noticing color is unaffected by the past history of changing correlations between the reinforced responses and the unnoticed color values. In contrast to the equality prediction above, if any one of the three assumptions is wrong, the Reversal *S*s should make more informed errors than do the Control *S*s.

Method

Subjects.—The *S*s were 33 paid volunteers from elementary history and psychology classes at Foothill Junior College who were randomly assigned to two groups (10 or 11 males and 6 females each).

Procedure.—The instructions were the same as those used in Exp. I and II. The classificatory responses were VEK and CEJ. The learning criterion was 10 successive correct responses.

Stimuli.—The patterns were identical to those used in Exp. II with the exception that the area which was colored within each figure was kept constant. Thus, there were one relevant and four irrelevant binary dimensions. Color was the relevant dimension for both groups.

Design.—The Control group of 16 *S*s learned a problem with fixed S-R assignments throughout. For 8 of these *S*s, the assignments were Red-VEK and Blue-CEJ; the other 8 *S*s had the opposite pairings. The Reversal group of 17 *S*s learned the same color-relevant problem but the response assignments were reversed on every second error that each *S* committed. On alternate errors, *S*'s response was confirmed (called "Correct") in accord with the instantaneous reversal of the assignments which *E* made as soon as *S*'s second error of a subseries occurred. The procedure was discussed above in connection with Table 3. By this procedure, it would not be feasible to reverse *S* on every error since he would always be told "Correct" and *E* would forfeit any control over what *S* learns.

Results

All but two *S*s, one in each group, met the learning criterion. The two nonsolvers arrived late for their experimental session and had less time than the other *S*s to complete the problem; therefore, they are excluded from the following analyses. Since both *S*s made about the same number of errors, their exclusion does not affect the comparisons.

The remaining 16 *S*s in the Reversal group averaged 7.00 reversal shifts before meeting criterion. The average numbers of informed errors were nearly equal for the two groups. For the Reversal group, the average number of informed errors was 7.81; for the Control group, it was 8.00. The *SD* of errors for the Control group was 8.22. Thus, the difference of .19 informed errors is not significant.

Two *S*s in each group learned after only one error. As a result, two *S*s in the Reversal condition were never reversed because they learned their initial response assignments. Removing these two *S*s from each group, the mean number of reversals was 8.00; the mean number of informed errors was 8.79 for the Reversal group and 9.08 for the Control group, a nonsignificant difference.

On those trials where a reversal occurred, *S* was told "Correct" when in fact he made an error. Such a procedure should serve to maintain an irrelevant hypothesis for at least one more trial. The net effect of this procedure would be to produce more "correct" responses before the last error for the Reversal *S*s than for the Controls. The mean numbers of corect responses prior to criterion for the Reversal and Control groups were 21.1 and 9.6, respectively, $t\ (29) = 1.99$, $p = .05$.

The mean trial of last error can be predicted for both groups once the mean errors for the Control group are known. These predictions are made to rule out the possibility that (*a*) length of success runs increased over trials in the Control group and (*b*) successive reversals tended to become more spaced out in the Reversal group over trials. For both predictions, the probability of a success prior to the last error is assumed to be constant and the a priori one-half. Let T_c be the total errors made by an *S* in the Control group; then his expected trial of last error is $2T_c$. The predicted mean trial of last error for the Control group was 16.00; the observed was 17.60. The difference was not significant by a matched *t* test, $t\ (14) = 1.75$, $p > .05$.

For the Reversal group, let T_r be the number of informed errors and r be the number of reversals before learning. Then the average trial of last error, n', for the Reversal group should be

$$n' = T_r + r + 1 + 2(T_r - 1). \quad [1]$$

The first two terms, T_r and r, in Equation 1 count the number of informed error trials plus the reversal trials. The additional terms $1 + 2(T_r - 1)$ are the expected number of correct responses for an *S* who makes T_r informed errors in the Reversal group. The T_r informed errors partition the successes as follows: there is an average of one success before the first error, and an average of two successes between each of the $T_r - 1$ remaining informed errors. By hypothesis, the informed errors should be the same for both groups, so that $T_c = T_r$. Secondly, r is related to T_r, for if *S* makes T_r informed errors, then his number of reversals should be $T_r - 1$, assuming that he makes one more error after his *r*th and final reversal. (Note that the average number of reversals was 7, or exactly $T_c - 1$.) Substituting into Equation 1 the relations $r = T_r - 1$ and $T_c = T_r$, the following relation is obtained between T_c and the average trial of last error for the Reversal group:

$$n' = 4T_c - 2. \quad [2]$$

Substituting the observed $T_c = 8.00$ into Equation 2, the predicted mean trial of last error (n') is 30.00. For the 16 solvers in the Reversal group, the observed value was 28.81; the *SD* was 26.09. The prediction is thus not significantly discrepant from the data.

The results of Exp. III favor a one-step, all-or-none interpretation of two-category concept identification in adult *S*s. In addition, the results indicate that the effective information promoting learning in these problems occurs on informed error trials. Finally, the results are consistent with the notion that *S*'s probability of solving after any given informed error is unaffected by the past history of inconsistent reinforcements to the relevant cue on which he solves.

Again the criticism may be lodged that our experiment had inadequate power to reject the null hypothesis. To provide further power for the test, we presently are running the design of Exp. III with larger groups of *S*s, a different problem, and more explicit instructions to *S* regarding the dimensions of the stimuli, the form of the solution, etc. To date, with 24 *S*s in each condition, the mean numbers of informed errors in the Control and Multiple Reversal groups are 8.22 and 7.94, respectively. Thus, the qualitative results of Exp. III are being replicated.

DISCUSSION

The Reversal and Control conditions in Exp. I and II resemble the standard ones used with this design on the continuity-noncontinuity issue. Judging from the review by Blum and Blum (1949), nearly all of the previous studies involved rats learning simultaneous discriminations with a small number of cues. On balance, that evidence favored a continuity position supplemented by constructs such as receptor orienting acts (e.g., Ehrenfreund, 1948). Whether such results should have a crucial bearing on a situational theory of adult human concept identification is a moot question. Writing for the continuity position, Spence (1940) pointed out early that the results from the animal studies may not be directly relevant to adult human learning mediated by complex symbolic mechanisms. Such mechanisms evidently are used by adults in solving concept problems, and current theorizing emphasizes such mechanisms (e.g., Bower & Trabasso, 1963a; Hunt, 1962; Kendler & Kendler, 1962; Underwood & Richardson, 1956). Our working hypothesis is that the extent to which an *S*'s discrimination learning fits the all-or-none as opposed to the incremental description depends on the extent to which symbolic mediating responses are available to *S*.

It would appear that one reason why the all-or-nothing model predicts accurately in these experiments is that the conditions promote "focus sampling" (Bruner, Goodnow, & Austin, 1956) because the memory load on *S* is otherwise overwhelming. The random-cue selection postulate implies that *S*'s selection following an error of a new focus sample of cues to test is not affected by the past history of response assignments for the various cues. Such random selection of a sample focus is reasonable only if *S*'s memory of specific past information is in some way impoverished. The experimental conditions presumably responsible for such poor memory include (*a*) the complexity of the stimuli, here 5 or 6 bits plus the 1-bit response, (*b*) the relatively rapid rate of presentation of this information (average time viewing each card was approximately 6 sec.), and (*c*) *S* has a specific set to identify the relevant cue, not to memorize and later recall the information he is seeing. In other experiments by us, direct tests of recall of specific information under these conditions showed the memory for six-card series to be very poor. Judging from the limited capacity of *S*s for quickly processing and storing such large amounts of information, it is not surprising to find that they resort to focus sampling of specific cues to test.

The present results extend previous findings (Trabasso, 1963) that single-cue concept problems can be characterized as a one-step learning process. However, it is clear that not all varieties of concept learning can be so simply described. Our aim was to explore initially the most elementary form of concept learning, in a situation similar to a conventional discrimination learning procedure. Obviously, the simple all-or-nothing model must be elaborated and extended before it will account for learning of compounds of simpler concepts (e.g., conjunctions or disjunctions of several cues). Such extensions are currently under investigation (Trabasso & Bower, in press).

REFERENCES

BLUM, R. A., & BLUM, J. S. Factual issues in the "continuity controversy." *Psychol. Rev.*, 1949, **56**, 33–50.

BINDER, A. Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.*, 1963, **70**, 107–115.

BOWER, G. H., & TRABASSO, T. R. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford Univer. Press, 1963, in press. (a)

BOWER, G. H., & TRABASSO, T. R. Working paper on concept identification. Unpublished manuscript, Stanford University Library, 1963. (b)

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, A. *A study of thinking.* Wiley: New York, 1956.

EHRENFREUND, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. physiol. Psychol.*, 1948, **41**, 408–422.

GRANT, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol. Rev.*, 1962, **69**, 54–61.

HAYES, K. J. The backward curve: A method for the study of learning. *Psychol. Rev.*, 1953, **60**, 269–275.

HUNT, E. B. *Concept learning.* New York: Wiley, 1962.

KENDLER, H. H., GLUCKSBERG, S., & KESTON, R. Perception and mediation in concept learning. *J. exp. Psychol.*, 1961, **61**, 1–16.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

KRECHEVSKY, I. A study of the continuity of the problem-solving process. *Psychol. Rev.*, 1938, **45**, 107–133.

MCCULLOCH, T. L., & PRATT, J. G. A study of the pre-solution period in weight discrimination by white rats. *Psychol. Rev.*, 1934, **18**, 271–290.

RESTLE, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, **69**, 329–343.

SPENCE, K. W. Continuous versus noncontinuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **54**, 223–229.

TRABASSO, T. R. Stimulus emphasis and all-or-none learning in concept identification. *J. exp. Psychol.*, 1963, **65**, 398–406.

TRABASSO, T. R., & BOWER, G. H. Component learning in the four-category problem. *J. math. Psychol.*, in press.

UNDERWOOD, B. J., & RICHARDSON, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51**, 229–238.

(Received November 23, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 419–421

SUPPLEMENTARY REPORTS

ROLE OF POSITIONAL CUES IN SERIAL ROTE LEARNING

WILMA A. WINNICK AND RHEA L. DORNBUSH

Queens College of the City University of New York

3 groups of Ss learned an identical list of nonsense syllables under 3 conditions of presentation: 1 group learned with a constant starting position; in a 2nd, starting position shifted by 1 syllable on successive trials; and in a 3rd, the starting syllable was varied randomly from trial to trial. ½ of the Ss in each group were given instructions indicating the possibility of a shift in starting position. Significant differences were found among the trials to criterion of the 3 noninstructed groups, supporting Ebenholtz' (1963) conclusion that positional cues play a role in serial learning. The instructed groups showed convergence of their learning curves compared to the noninstructed.

Ebenholtz (1963) demonstrated that eliminating position as a basis for responding in serial learning brought about significant decrements in the rate of learning. This was shown by comparing the trials required by two groups to learn an identical sequence of nonsense syllables; the control group learned with a constant starting position, while for the experimental group the syllable which appeared in the starting position was changed from trial to trial in a prearranged manner.

The present experiment, independently planned, used the same control group and one experimental group similar to Ebenholtz. In an additional group it was possible for *S* to predict subsequent starting positions; it was expected that such a group would learn faster than the first experimental group but not so fast as the control. In addition to the standard anticipation instructions, half of each group was given information indicating the possibility of a shift in starting position.

Method.—All *S*s learned the same list of 14 nonsense syllables arranged in identical serial order in a circular series on the memory drum. In the control group (Group 1) starting position was the same on all trials for any one *S*; each *S* started with a different syllable to control for possible differential difficulty of items in different serial positions. In the first experimental group (Group 2) the starting syllable was shifted on succeeding trials in a predictable manner to the next in position; thus, the second syllable on any one trial became the first on the next trial. In the second experimental group (Group 3), starting position was shifted in random manner from trial to trial. For all groups, the immediately preceding nonsense syllable served to cue anticipation of the first nonsense syllable on any trial; for Group 1, this was also the last nonsense syllable on the list. For half the *S*s in each of the groups (Subgroups 1A, 2A, 3A) instructions were the standard anticipation instructions. For the other half (1B, 2B, 3B), the following information was added: "Just try to learn to associate one syllable with the next, as the starting position may shift on any one trial." All *S*s were instructed to pronounce the syllables, which appeared in the memory drum at a 2-sec. rate and with a 14-sec. intertrial interval.

Sixty undergraduates without experience in this type of learning served as *S*s; they were randomly assigned to form six experimental groups with 10 *S*s in each group. In addition, 6 *S*s had to be dismissed when it became apparent that they were excessively upset by their inability to learn the materials: one each from Groups 2A and 1B and two each from 2B and 3B.

Results and discussion.—Table 1 shows the mean number of trials to the criterion of two

TABLE 1

MEAN TRIALS TO CRITERION, MEAN ERRORS PER TRIAL, AND MEAN INTRUSIONS PER TRIAL

	Group 1		Group 2		Group 3	
	A	B	A	B	A	B
Mean trials to criterion	11.9	14.9	16.0	16.1	20.0	17.8
SD	2.9	5.1	3.9	6.3	5.4	7.0
Mean total errors/trial	4.4	4.6	4.9	4.5	5.0	4.9
SD	.8	.8	.8	.8	.8	1.5
Mean total intrusions/trial	1.7	1.4	1.6	1.9	1.7	1.7
SD	.5	.2	.6	1.6	.8	.2

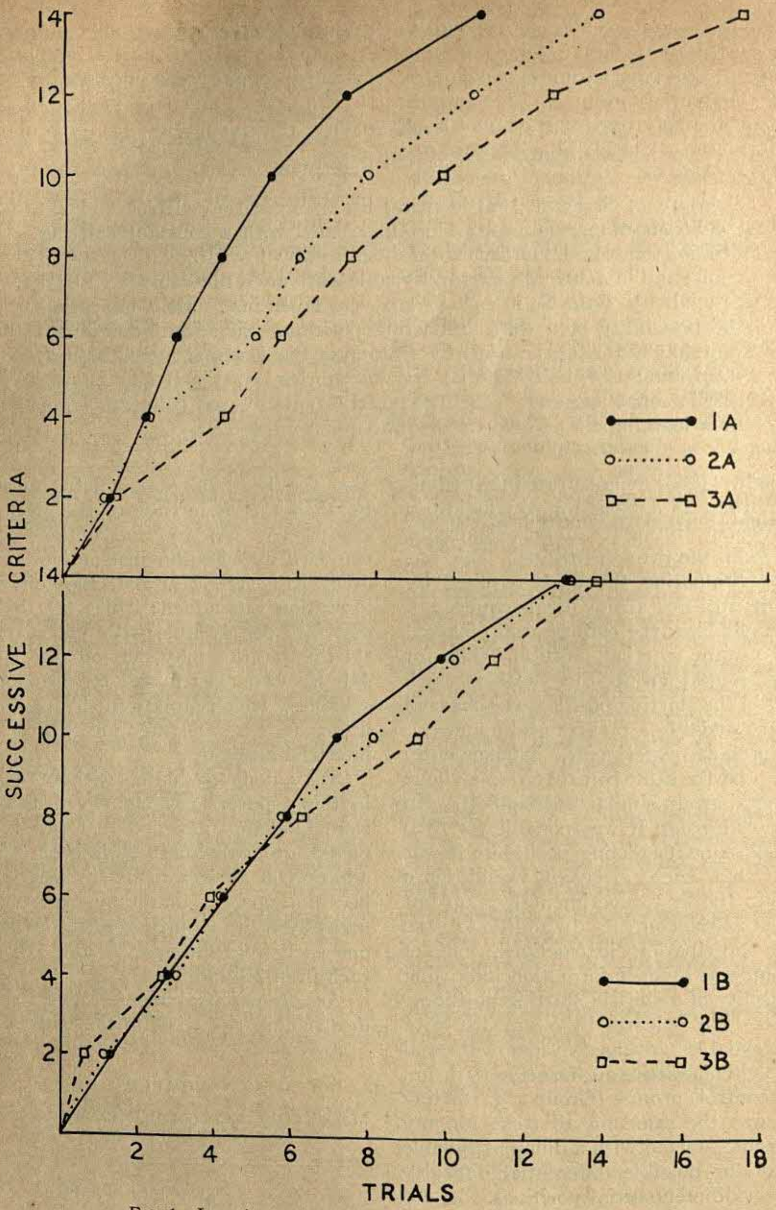


FIG. 1. Learning curves for the six experimental conditions.

errorless trials and the mean number of errors and intrusions per trial for the six groups.

The two-way analysis of variance, called for by the experimental design, compared trials to criterion and found only the interaction between instructions and method to be significant, $F(2, 54) = 5.9$, $p < .01$. Subsequent t tests indicated that Group 1B showed a significant decrement in learning with the addition of instructions, compared to Group 1A; for 1B, instructions provided misinformation. For Group 3B (randomly varying position) the addition of instructions produced a slight but not significant improvement compared to 3A. Comparisons of Groups 2A and 2B showed the instructions to have no effect at all, apparently because they provided no information other than that which *S* could soon provide himself.

In order to compare these data with those of Ebenholtz, whose results correspond to those of the present noninstructed groups, a one-way analysis of variance was undertaken for the data of Groups 1A, 2A, and 3A, yielding an among-groups F (2, 27) of 8.32, significant at the .01 level. The Duncan test revealed comparisons of Groups 1 and 3 to be significant at the .01 and 2, 3 and 1, 2 at the .05 level of significance. Changing the starting syllable from trial to trial, which effectively deprives *S*s of a positional cue, has then significantly affected trials to learn. Furthermore, where *S* could predict the starting position on subsequent trials (Group 2), trials to learn improved over what they were when this prediction could not be made, but were poorer than with a constant starting position.

Figure 1 shows the number of trials required to reach successive levels of performance for the three noninstructed groups and for the three instructed groups. Without instructions, the curves of the three groups are disparate, and curves for 1A and 3A show strong resemblance to those of Ebenholtz' control and experimental groups. The convergence of performance of the three groups induced by the instructions is apparent from the curves of 1B, 2B, and 3B. The fact that the instructions could effect an upset of 1B's learning and an improvement in 3B's is considered substantiation of Ebenholtz' conclusion that positional cues play a role in serial verbal learning. The differential effect of instructions is further taken to indicate that it was lack of positional cues that upset the learning of the noninstructed groups, not the breakdown of the associative structure of the list which might occur in the two experimental procedures used. Specifically, when *S*s were told that position would change (as in 3B), they were somewhat better than when position shifted without instructions (as in 3A). On the other hand, when *S*s in 1B were given such instructions, but position was not changed, they were worse than the uninstructed group 1A.

REFERENCE

EBENHOLTZ, S. M. Serial learning: Position learning and sequential associations. *J. exp. Psychol.*, 1963, 66, 353-362.

(Early publication received May 13, 1963)

Journal of Experimental Psychology
1963, Vol. 66, No. 4, 421-422

NOVELTY EFFECTS IN CUE ACQUISITION AND UTILIZATION [1]

LOY S. BRALEY AND DONALD MICHAEL JOHNSON

University of California, Santa Barbara

A previous study had shown that irrelevant cues in the 2nd of a 3-stage concept learning task impaired utilization of these cues when they became relevant in Stage III, as compared to controls with no cue exposure (NCE). As a replication, and to determine if incremental cue extinction occurs, Ss were exposed to 4, 10, or 16 trials of irrelevant cue exposure (CE) in Stage II. Also, 15 Ss were run on Stage III alone (III A) to determine if differences were due to NCE facilitation or CE impairment. Results were: (a) successful replication, (b) Stage III performance unrelated to amount of prior cue exposure, (c) no difference between CE and III A groups, indicating NCE facilitation due to differential cue saliency.

In a previous paper Braley (1962) presented results to suggest that in a concept learning task the introduction of additional (but irrelevant) cues after an initial discrimination has been learned functioned so as to inhibit, rather than facilitate, the utilization of these cues when they later became the basis for correct discrimination. The experimental task was three stages with simultaneous presentation of a positive and negative instance on each trial. The *S* was required to learn a concept to criterion in Stage I. Stage II was continuous with Stage I and entailed presentation to experimental *S*s (CE—cue exposure) of additional (but irrelevant) cues on the positive instances, while control *S*s (NCE—no cue exposure) were simply continued beyond criterion for an equivalent number of trials. Stage III was also continuous with Stage II and required both groups to learn a new concept defined by those cues that the CE group had been exposed to in Stage II.

---

[1] The junior author's collaboration in this study was made possible by the National Science Foundation's Undergraduate Research Participation Program.

TABLE 1

MEANS AND *SDs* OF TRANSFORMED TRIALS TO CRITERION FOR CONCEPT LEARNING: STAGE III

Number of Trials in Stage II	Irrelevant Cues Present in Stage II		Irrelevant Cues Absent in Stage II	
	M	*SD*	*M*	*SD*
4	5.12	3.13	2.34	1.19
10	5.08	2.98	2.79	1.64
16	5.53	2.47	2.67	2.28

An explanation of the rather surprising results, that the NCE *S*s did significantly better on Stage III learning, was offered in the original paper in terms of stimulus novelty. That is, the new cues in the Stage III task constituted a novelty condition for the NCE groups while these same, now relevant, cues represented the familiar for CE *S*s and hence failed to generate appropriate arousal or to mobilize the "orienting reflex" (Luria, 1957).

At least two questions occur in connection with these results: (*a*) Is the more rapid learning for *S*s not previously exposed to the irrelevant cues an artifactual consequence of using only a 10-trial cue exposure condition in Stage II? Further, if some process of conditioning or extinguishing of cues is occurring in Stage II then varying the number of irrelevant CE trials in this stage should produce differential effects on Stage III learning. (*b*) The second question has to do with the author's implication that the poorer performance of the CE groups in Stage III learning reflects some degree of inhibition. Rather, it might be argued that prior cue exposure has no effect and the difference between groups is due solely to the facilitation of performance for the NCE groups as a consequence of stimulus novelty increasing the saliency of the relevant cues.

Method.—To answer the first question, whereas only 10 CE trials were employed in the original study, this investigation involved separate groups receiving 4, 10, or 16 exposures to the irrelevant cues in Stage II. Three control groups were simply continued beyond criterion for 4, 10, or 16 trials. Of 60 *S*s from a volunteer university summer school population 10 *S*s were randomly assigned to each of the 6 groups. With the exception of the elimination of the Anxiety and Primary Cue variables, all other details of apparatus, stimulus materials, instructions, and procedure were as in the previous study.

As an approximation to answering the second question an independent group of 15 *S*s from the above population was run only on the Stage III concept learning task entirely omitting Stage I and II performance.

Results and discussion.—An analysis of variance, on square root transformed trials to criterion, of varying the number of CE trials yielded a significant $F\ (1, 54) = 19.85$ only for the main effect; viz., more rapid learning for NCE groups. This constitutes a confirmation of the previous results and suggests they are not an artifactual consequence of number of CE trials employed. Also the analysis of variance indicates, confirmed by inspection of means in Table 1, no evidence for differential acquisition of cues as a function of the number of CE trials. If, as Luria (1957) and Sokolov (1954) have maintained, orienting responses occur to new or changed stimuli it appears that the novelty effect of these stimuli undergoes extinction in 4 or fewer trials in light of the similar means for 4-, 10-, and 16-trial groups. In like manner the probability of use of these irrelevant cues in Stage III learning seems unrelated to the frequency of previous occurrence.

A test of whether the difference between CE and NCE groups reflects inhibition (extinction or adaptation of cues) for the CE group or facilitation for the NCE group was made by *t* testing the difference between the mean for combined CE groups and the mean of the independent group of 15 *S*s ($M = 5.62$, $SD = 2.38$) run only on the Stage III learning task. Significance was not obtained indicating that the difference between CE and NCE conditions was basically a matter of facilitation for the latter.

The conclusion that seems apparent from these results is that in a three-stage concept learning task such as this, nonreinforcement signals the change to Stage III and awareness of change demands the generation of new hypotheses. New, or changed, cues introduced at this point are likely to be the focus of these hypotheses and hence facilitate utilization. The prior occurrence on the positive instances in Stage II of the now relevant cues for 4, 10, or 16 trials seems to have had no effect on the probability of their use in Stage III concept learning.

REFERENCES

BRALEY, L. S. Some conditions influencing the acquisition and utilization of cues. *J. exp Psychol.*, 1962, 64, 62–66.

LURIA, A. R. The role of language in the formation of temporary connections. In B. Simon (Ed.), *Psychology in the Soviet Union.* Stanford: Stanford Univer. Press, 1957. Pp. 115–129.

SOKOLOV, E. N. Higher nervous energy and the problem of perception. Paper read at Fourteenth International Congress of Psychology, Montreal, Canada, 1954.

(Received October 26, 1962)

Journal of
Experimental Psychology

VOL. 66, NO. 5 NOVEMBER 1963

EFFECTS OF LABELS ON CHILDREN'S PERCEPTION AND DISCRIMINATION LEARNING[1]

PHYLLIS A. KATZ[2]

Yale University

The present study tested the hypothesis that the nature of verbal labels associated with stimuli influences the subsequent perception of those stimuli. 48 children were given differential verbal training to associate nonsense syllables with 4 highly similar visual forms. One group learned common labels; another, distinctive labels; and a 3rd, no labels. Following verbal training, perceptual and discrimination learning tasks were administered to all Ss. Children who had associated common, as opposed to distinctive, labels to 2 stimuli perceived the stimuli as identical significantly more often and exhibited greater difficulty in learning to discriminate between them.

The importance of verbal labels in the discrimination process has been emphasized by Dollard and Miller (1950). These investigators have suggested that when labels become associated with physical objects, the linguistic cues will influence the generalization gradients of the stimuli, such that: (*a*) distinctive labels will reduce the initial generalization tendencies, whereas (*b*) similar or identical labels will increase the amount of stimulus generalization.

This hypothesis has been confirmed in a number of recent studies which have demonstrated that linguistic cues associated with stimuli influence the transfer of instrumental discriminatory responses to those stimuli (e.g., Goss, 1953; Jeffrey, 1953; McAllister, 1953; Rossman & Goss, 1951; Spiker, 1956). However, other investigations which have attempted to demonstrate the effects of labels on perceptual judgments of same or different have culminated in negative findings (e.g., Arnoult, 1953; DeRiviera, 1959; Robinson, 1955; Smith & Jones, 1960).

The ambiguity of these findings is puzzling. If labels affect the perceived similarity of stimuli, and if motoric discriminatory responses and verbal judgments of stimulus similarity are both considered to be measures of this underlying perceptual

---

[1] This study is based upon a dissertation submitted to Yale University in partial fulfillment of the requirements for the PhD degree. The writer would like to express her appreciation to Edward Zigler for his helpful guidance.

This research was supported by funds from Grant M-1849, National Institutes of Health, directed by Lyle V. Jones.

[2] Now at New York University.

process, the question can be raised as to why the two measures should yield different results.

It should be noted that this question is difficult to answer on the basis of existing literature since the above cited studies employed either one or the other measure, under differing experimental conditions with diverse *S* populations. The present experiment represents an attempt to explore this question further, employing a design in which both instrumental discriminatory responses and verbal judgments of similarity were made to the same set of stimuli by the same *S*s.

It was hypothesized that with children, both measures would reflect the influence of labels, such that: (*a*) the addition of a common label to two similar stimuli would result in increased perceptual equivalence, whereas (*b*) the addition of differential labels to these same stimuli would result in increased perceptual distinctiveness. Piaget's (1951) work on nominal realism suggests that labels may play a greater role in the discriminal processes of young children than those of older children. Thus, it was further predicted that the effects of labels upon perception would be more pronounced in younger, than in older children.

METHOD

Design

The study was conducted in three stages. The initial stage consisted of verbal training in which nonsense syllables were associated with four visual forms. Three experimental conditions were introduced: a *common-label* (CL) group was taught to associate two syllables with the four figures, i.e., two figures per label; a *distinctive label* (DL) group was taught to associate a different syllable with each of the four figures; finally, a *no-label* (NL) group was shown the figures without labels. Half of each group was composed of 7-yr.-olds and half of 9-yr.-olds.

In the second stage, all groups were given a perceptual task which involved making judgments of "same" or "different" to tachistoscopically-presented pairs of the geometric figures. In the third stage, a discrimination learning task was administered to all *S*s.

Subjects

The *S*s were 48 children, 24 of each sex from a public school in New Haven, Connecticut. Half were chosen from the first and second grades, and half from the fourth grade. The mean chronological ages of the younger and older groups were 7.0 and 9.6 yr., respectively. These groups were chosen to correspond to the age levels at which the first and third stages of nominal realism are reported to occur (Piaget, 1951). Within each age group, *S*s were randomly assigned to experimental conditions, with the restriction that both sexes be equally represented in each group.

Stimuli

The visual stimuli were four geometric forms designed according to Attneave's (1954) system for the construction of random forms. The stimuli were constructed from Masonite and differed in shape, but were all the same color, (maroon,) and approximate size (3½ in. square). An illustration of the forms used is presented in Fig. 1.

The verbal labels were chosen from nonsense syllable lists of 95% association value (Hilgard, 1951), and were: BUZ, RIC, JAN, and SOL. The syllables were randomly assigned to the figures for each *S*. For *S*s in Group CL, two syllables were randomly selected from the four.

Apparatus

During the verbal learning and perceptual tasks, colored slides of the four forms were projected by a Revere 888 slide projector onto a 40-in. white square screen from a distance of 7½ ft. At this distance, the figures appeared their actual size. A Graflex SVE tachistoscopic shutter was employed in the perceptual judgment task at a constant shutter speed of .2 sec. with a medium brightness opening of 5.6.

At the start of the perceptual judgment task, *S* was presented with two telegraph keys mounted on wooden blocks. One key was marked "same"; the other "different." The depression of either of the two telegraph keys illuminated a small red or green light

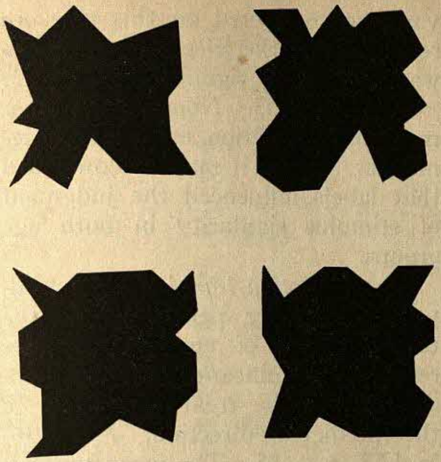


FIG. 1. Stimulus figures employed.

on a control board, thus informing *E* of *S*'s choice. The apparatus employed for the discrimination task consisted of a horizontal board with a collapsible Masonite vertical screen. During this phase, the actual forms were used, mounted on unpainted blocks of wood 4 in. square.

Procedure

Upon entering the experimental room, *S* was shown three boxes of toys. One contained balloons, the second, small animals, and the third, more expensive toys, e.g., jacks, modeling clay, guns, marbles, etc. The *S*s were instructed that they were going to play some games and could win the "good" toys if they played well.

Verbal training.—The slides of the four nonsense forms were introduced to the children as "pictures of funny-looking animals from Mars," and they were asked to pretend that they were space explorers. Each of the forms was initially shown twice to all *S*s, with the appropriate labels given to *S*s in Group CL and Group DL. The four figures were then presented individually for 2 sec. in a predetermined random order of 150 trials. The *S*s in the label groups were instructed to look carefully and tell the name of each figure. A correction procedure was used. In order to minimize the possibility that control *S*s would supply their own labels, they were asked to count the figures as they appeared. Verbal support was given to all *S*s every tenth trial, i.e., "You're doing fine," "You really know how to play this game," etc.

Perceptual task.—Immediately following the 150 verbal training trials, the telegraph keys were introduced, and *S* was instructed:

> This time you'll see two animals on each picture and they'll come on much faster. Sometimes both animals will be twins and look alike. Sometimes they'll be different. When I show you these pictures, I want you to decide whether the two animals on the picture are the same or different. If they're alike, like twins, you press this key that says "same." If they're not alike, you press this other key that's marked "different."

A practice slide containing two identical triangles was used to illustrate the concept of "same"; a triangle and a square were used to illustrate "different." Following eight trials with the practice slides, all *S*s were shown 28 pairs of the nonsense figures. Each slide was presented for 2 sec. with an intertrial interval of 10 sec. during which time *S* made a judgment of "same" or "different." The 28 slides contained: (*a*) 8 pairs of identical forms, and (*b*) 20 pairs of two different forms. Ten slides in this latter group contained pairs of different figures previously associated with a common label for Group CL. The actual forms used were matched across groups.

Discrimination task.—Following the perceptual judgment task, three discriminanda were introduced, and *S* was instructed:

> One of the animals has a paper goldfish like this under it. You pick up the one you think has the fish under it. If you're right, you can keep the fish and at the end, you can trade them in for prizes. You can only lift one at a time. Let's see how many fish you can win.

The reinforced stimulus was always one of the two forms previously associated with a common label for *S*s in Group CL. The criterion was either 5 consecutive correct choices or 50 trials. The *E* rearranged the position of the forms after each trial behind a vertical screen.

At the end of the experiment, all *S*s were permitted to choose one of the more expensive toys. Although total testing time ran as long as 45 min. for some children, it appeared as if the rewards and the variety of tasks were sufficient to maintain the children's attention and interest.

RESULTS

Verbal learning.—The measure employed for comparing Group CL and

Group DL on verbal learning was the number of correct responses per block of 50 trials. Group CL, however, would be expected to make more correct responses by chance alone than Group DL. Therefore, the number of correct responses was corrected for chance with the formula $R-W/n-1)$, where R is the number of correct responses, W the number of wrong responses, and n the number of distinct response categories available (Smith & Jones, 1960). The mean corrected scores on the last block of trials for Group CL and Group DL were 43.7 and 47.1, respectively. An analysis of variance performed on the corrected scores indicated that although both groups improved over blocks of trials, $F(2, 95) = 107.77$, $p < .001$, the difference associated with treatments was not statistically significant.

Perceptual judgments.—An analysis of variance performed on the number of "same" judgments made to the eight identical pairs of figures yielded no significant main or interaction effects. Thus, it may be concluded that differences in verbal training did not affect the judgments of pairs that were actually identical.

The measure most relevant to the hypothesis initially advanced is the number of "same" judgments *S*s made to the 10 pairs of nonidentical figures which were previously associated with varying types of labels. For Group CL, both figures in each pair were associated with a common label; for Group DL, a distinctive label was attached to each figure; and for Group NL, no labels were associated with the forms. (The physical forms were the same for all groups.) The mean number of "same" judgments made to these 10 pairs of figures by Groups CL, NL, and DL was 4.31, 3.06, and 1.94, respectively. An analysis of variance conducted on this measure revealed that the effect of treatments was statistically significant, $F(2, 42) = 4.70$, $p < .05$. None of the other main or interaction effects was significant. Thus, it may be concluded that labels influenced the judgment of stimulus similarity in both age groups.

Discrimination learning.—An analysis of variance performed on the total number of correct responses revealed a significant difference associated with the treatment effect in the predicted direction, $F(2, 36) = 4.55$, $p < .05$. The mean number of correct responses for Groups CL, NL, and DL was 16.31, 21.25, and 25.92, respectively. The prediction that there would be a significant Age×Treatment interaction received no support.

In addition to the number of correct responses, the kinds of responses the groups made following a reinforced trial were subjected to analysis. It should be recalled that *S*s in Group CL had previously associated a common label to two of the three discriminanda. The position initially advanced suggests that Group CL would exhibit perceptual confusion with respect to the two forms associated with the same name. It would be expected, therefore, that following a correct response, *S*s in Group CL would tend to choose either the reinforced form or the one associated with the same name. Figure 2 presents the data relevant to this hypothesis.

The analysis of variance conducted on this measure supports the hypothesis. It can be seen in Fig. 2 that the most frequently chosen figure following a reinforced trial is the reinforced figure, $F(1, 74) = 13.29$, $p < .001$. The *S*s in Group CL, however, tended to choose the nonreinforced figure associated with the same name as

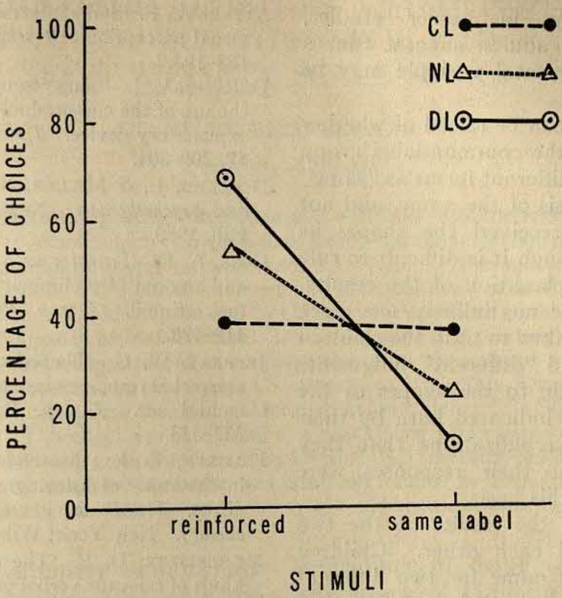


FIG. 2. Percentages of stimulus choices on trials following reinforcement.

frequently as the reinforced figure. This Figure × Treatment interaction is statistically significant at the .05 level.

DISCUSSION

The most significant finding of the present study was that differences in verbal training resulted in differences in performance on subsequently administered perceptual judgment and discrimination learning tasks. Children who had been taught to associate a distinctive label to visual forms more readily judged these forms as different and more rapidly learned a discrimination employing them than did children who had a common label associated with these same forms.

These results are in accordance with Dollard and Miller's hypotheses concerning the acquired equivalence and distinctiveness of cues. In addition, the findings related to the perceptual judgment task might also be interpreted as support for Whorf's (1956) principle of the linguistic determinism of perception. These latter findings stand in opposition, however, to the negative results obtained in the previously cited studies which have attempted to demonstrate the influence of language on perceptual judgment. This raises the question of why the present findings are discrepant with those of earlier studies employing a similar kind of measure.

The method of the present investigation differed from that previously employed in that tachistoscopic exposures of very similar stimuli were utilized, and children rather than adults were used as *S*s. The use of rapid exposures would decrease the perceived magnitude of existing differences in the stimulus, and, as Piaget suggests, the use of children should result in a greater dependence upon labels. Thus, it can be argued that the present study employed more optimal conditions for demonstrating linguistic effects on perception than did the earlier studies.

It should be noted that the prediction that more pronounced effects would occur in younger children was not confirmed. Despite the failure to confirm this specific prediction, the positive findings of the present study, which employed children, as opposed to the negative findings of earlier studies, which employed adults, suggest that a general developmental variable may be operative.

The question can be raised of whether the children in the common-label group judged the two different forms as "same" solely on the basis of the name, and not because they preceived the shapes as identical. Although it is difficult to rule out such an explanation of the results, it appears to be an unlikely one since great care was taken to train the children that "same" and "different" judgments were to be made to the shapes of the stimuli. All *S*s indicated both by their behavior and verbalizations that they understood that their responses were to be made on this basis.

As predicted, the results on the two tasks paralleled each other. Children with a common name for two different stimuli not only judged them as identical more often, but also exhibited perceptual confusion in learning instrumental discriminatory responses to the stimuli. This was clearly reflected in the finding that following reinforcement, the group given common labels chose the previously reinforced figure and the incorrect figure associated with the same name with an equal frequency. The group with distinctive labels, on the other hand, chose the reinforced figure with a much greater frequency. Thus, the overall findings of this study may be taken to indicate that labels influence the perceptual behavior underlying both the judgment of stimulus similarity and discrimination learning.

REFERENCES

ARNOULT, M. D. Transfer of predifferentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45**, 401–409.

ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.

DERIVIERA, J. Some conditions governing the use of the cue-producing response as an explanatory device. *J. exp. Psychol.*, 1959, **57**, 209–304.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

GOSS, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46**, 419–428.

JEFFREY, W. E. The effects of verbal and nonverbal responses in mediating an instrumental act. *J. exp. Psychol.*, 1953, **45**, 327–333.

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.

MCALLISTER, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329–336.

PIAGET, J. *The child's conception of the world.* New York: Humanities Press, 1951.

ROBINSON, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–114.

ROSSMAN, I. L., & GOSS, A. E. The acquired distinctiveness of cues: The role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.

SMITH, T. A., & JONES, L. V. Effects upon verbal learning of stimulus similarity, response uncertainty, and concept formation. Unpublished manuscript, University of North Carolina, 1960.

SPIKER, C. C. Stimulus pretraining and subsequent performance in the delayed reaction experiment. *J. exp. Psychol.*, 1956, **52**, 107–111.

WHORF, B. L. *Language, thought and reality.* New York: Wiley, 1956.

(Received December 5, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 429–438

EYELID CONDITIONING AS A FUNCTION OF CS INTENSITY, UCS INTENSITY, AND MANIFEST ANXIETY SCALE SCORE[1]

SALLY BELL BECK[2]

University of Illinois

The effects and possible interaction of 3 variables—CS intensity, UCS intensity, emotionality of S (MA scale score)—on the formation of both the conditioned eyelid response and the "voluntary" response were investigated. All 3 variables were found to be positively related to CR formation. The effect produced by CS intensity was much greater than that obtained in previous studies. It was suggested that this heightened effect could be interpreted as an adaptation level phenomenon or contrast effect since, in this study, CS intensity was a within-S variable. The only interaction observed was for percent long latency CRs (300–500 msec). This effect was attributed to a near ceiling performance of the High MA scale groups under strong CS and UCS intensity conditions. Voluntary response formation was positively related to both CS and UCS intensity but not to emotionality of S (MA scale score).

The role of intensity of both the conditioned stimulus (CS) and the unconditioned stimulus (UCS) has received considerable attention in studies of classical conditioning. According to Hull (1949), stimulus intensity dynamism combines multiplicatively with such variables as drive and habit strength to determine reaction potential.

In general, (Passey, 1948; Spence, 1953) a direct positive relationship is reported to exist between UCS intensity and the amount of eyelid conditioning. Although UCS intensity has been regarded primarily as a determiner of drive (*D*) within the framework of Hullian theory, recent work by Spence and his associates (Ross & Hunter, 1959; Spence, 1953; Spence, Haggard, & Ross, 1958a, 1958b; Trapold & Spence, 1960) has suggested that it may be a determiner of habit strength as well.

The relationship between CS intensity and the conditioned response (CR), however, presents a rather conflicting picture. Pavlov (1927) had indicated that unless a strong UCS was utilized, an ordinarily suitable CS might prove ineffective. In addition, Razran (1957) demonstrated that the magnitude of the unconditioned response (UCR), which is in part a function of UCS intensity, determines the effect of CS intensity on response strength. When Walker (1960) noted that in general a positive relationship between CS intensity and response strength was found to exist for data obtained during acquisition (when the UCS is present) and little or no relationship for data obtained during extinction (with the UCS absent), she turned to an explanation based on possible CS-UCS interaction. While no significant CS-UCS interaction effect was obtained by Walker (1960) for the total conditioned response data, a trend toward CS-UCS interaction for the earlier stage of conditioning was reported.

---

[1] This article is based on a dissertation submitted to the graduate college, University of Illinois in partial fulfillment of the requirements for the PhD degree. The author wishes to especially acknowledge her chairman, G. Robert Grice, for his valuable help and advice throughout the course of the experiment and W. E. Kappauf, L. A. Hellmer, J. McV. Hunt, and L. Stolurow for their helpful suggestions. Thanks are also given to Jacque Goldman for hours of computational aid. This study was supported by Research Grants G-4416 and G-14223 from the National Science Foundation.

[2] Now at Purdue University, Indianapolis Campus.

In addition to studies of direct manipulation of UCS and CS intensity, a series of recent investigations (Spence, 1958) have directed attention to the selection of *S*s who differ in the intensity of their emotional responsiveness (Taylor, 1953). A number of studies (e.g., Spence, 1958; Taylor, 1951) which have selected *S*s on the basis of response to items on the Taylor Manifest Anxiety scale (*MA* scale) (Taylor, 1953), assumed to measure emotional responsiveness (Spence, 1958), report a significantly higher level of conditioning for high *MA* scale groups. In some studies, however, this relationship is not obtained (e.g., King, Kimble, Gorman, & King, 1961). Spence and Taylor (1953) and Baron and Conner (1960) have shown the overall relationship of *MA* scale score and level of conditioning to be positive and monotonic.

The results from a study (Spence, 1956), however, in which high and low *MA* scale *S*s were conditioned with a very weak UCS and *two* levels of CS intensity indicate a possible interaction between CS intensity and level of emotionality for the latter stage of acquisition. High and low *MA* scale groups were found to differ significantly in level of response under conditions of a strong CS, but not under a weak one (where the high *MA* scale group gave fewer CRs than those of the low *MA* scale group). Furthermore, high and low CS intensity level groups differed significantly within the high *MA* scale group but not within the low *MA* scale group.

Since there is some evidence that high *MA* scale *S*s may respond differently to various levels of CS intensity, it is possible that *S*s selected on this basis may be more likely to show the CS-UCS interaction which Walker (1960) failed to obtain. The present study utilized a factorial design involving three variables: CS intensity, UCS intensity, and emotional responsiveness as measured by the *MA* scale. Groups of *S*s, selected as high or low in emotional responsiveness, were trained under combinations of strong or weak CS and UCS intensity levels. This was expected to increase the opportunity to observe differential responding on the basis of intensity and the presence of a CS-UCS intensity interaction if it does exist.

In addition, all anticipatory responses were differentiated on the basis of both form and latency in order to determine how these variables affect the formation of not only the conditioned response but of the "voluntary" response as well.

METHOD

Subjects and design.—The *S*s were 152 female students chosen from 392 women enrolled in an introductory psychology course at the University of Illinois. The *S*s were selected on the basis of extreme scores on a short form of the Biographical Inventory consisting of the 50 original *MA* scale anxiety items and 50 buffer items. Seventy-six high *MA* scale scorers were defined as highly emotionally responsive or high *MA* (scores of 23 and above) and 76 low scorers designated as less emotionally responsive or low *MA* (scores of 11 and below). These two groups of *S*s were divided and assigned at random to either a strong or weak UCS condition thus forming four groups of 38 *S*s each. All *S*s in the four groups (high *MA*-strong UCS; high *MA*-weak UCS; low *MA*-strong UCS; low *MA*-weak UCS) received 80 paired presentations of tone (CS) and puff (UCS). A loud tone (CS) was presented on 40 of the trials and a soft tone (CS) on the other 40 according to a prearranged irregular schedule. Five loud and five soft CS stimulus presentations were given within each block of 10 trials. Neither intensity occurred on more than two successive trials throughout the course of conditioning.

Apparatus.—Two intensities of a 1,000-cps tone were employed as the CS; a weak intensity of 30 db. and a strong intensity of 80 db. above an average threshold. The tone was generated by a Hewlett-Packard Model 205 AG audio oscillator and delivered to *S* by means of a set of Permoflux PDR-8 earphones. A low pass electrical filter was used to eliminate any audible clicks through the earphones and tone intensities were measured after the filter.

Two intensities of air puff were employed as the UCS; a weak intensity of .5 lb/sq in and a strong intensity of 5 lb/sq in. Appropriate pressures were assured by the use of a Matheson variable air pressure reducing valve. The CS duration was 500 msec., and its termination coincided with the onset of the UCS which then lasted 40 msec. Onset and offset of the stimuli were controlled by means of a Hunter electronic timer. Eyelid movement was recorded by a potentiometer polygraph system described by Spence (1953). The entire experimental session of 80 trials was automatically presented by use of programing equipment to insure greater precision and reduction of *E* error.

Conditioning procedure.—The *S*s were seated in the conditioning room which was soundproof with an incident illumination of 2.5/ft-c measured at *S*'s eyes. A buzzer ready signal was presented randomly at 2, 3, or 4 sec. preceding each trial. When the buzzer was sounded, *S* was instructed to look at a white disc (placed approximately 3 in. above eye level on the wall 60 in. directly in front of *S*'s chair) blink once, and continue to look at the disc until a tone (CS) sounded. Intertrial intervals of 15, 20, or 25 sec. were employed according to a prearranged random schedule with an average intertrial interval of 20 sec.

Measurement of response.—Anticipatory responses were scored in two ways: First, according to form (Spence, 1958). "Voluntary" responses were identified by their sharp, smooth closure which is maintained until after the onset of the UCS; while the CR's form is irregular. Second, according to latency. Short latency responses which occurred in the interval 150–300 msec. after CS onset and those with longer latencies from 300 to 500 msec. were tabulated separately. Twenty-five *S*s were discarded due to equipment failure.

RESULTS

The mean number of anticipatory responses, including responses of voluntary form, is shown in Fig. 1 plotted as a function of blocks of 10 trials. Within each block of 10 trials, 5 trials consisted of the loud tone CS and 5 trials of the soft tone CS. Anticipatory responses were defined as a 1-mm. or greater penwriter incline following a 5-mm. nonincline during the time interval from 150 to 500 msec. following CS onset. It may be noted that loud CS (tone) presentation led to a consistently higher level of performance than those receiving the weak intensity UCS. Since there is evidence (Taylor, 1954; Walker, 1960) that the production of short latency responses may be subject to different laws from those governing the longer latency responses, anticipatory responses were differentiated on the basis of both form and latency and separate analyses performed to evaluate the reliability of observed differences during acquisition.

There appears to be no entirely satisfactory method for dealing with the problem of voluntary response in a study such as this one. The procedure of elimination of "voluntary responders" appears to be of questionable validity here—since the frequency of voluntary response is related to the experimental variable, the remaining groups, after such elimination, could no longer be regarded as samples from the same population. If the problem is ignored and all responses are included, the frequencies in some groups are inflated more than others by the presence of voluntary responses. The decision made was to deal with percent CR on only those trials when voluntary responses did not occur. This procedure also raises problems, including possible variation in reliability among *S*s and potential contribution to the variance by variations in the denominator. However, in the present study at least, it

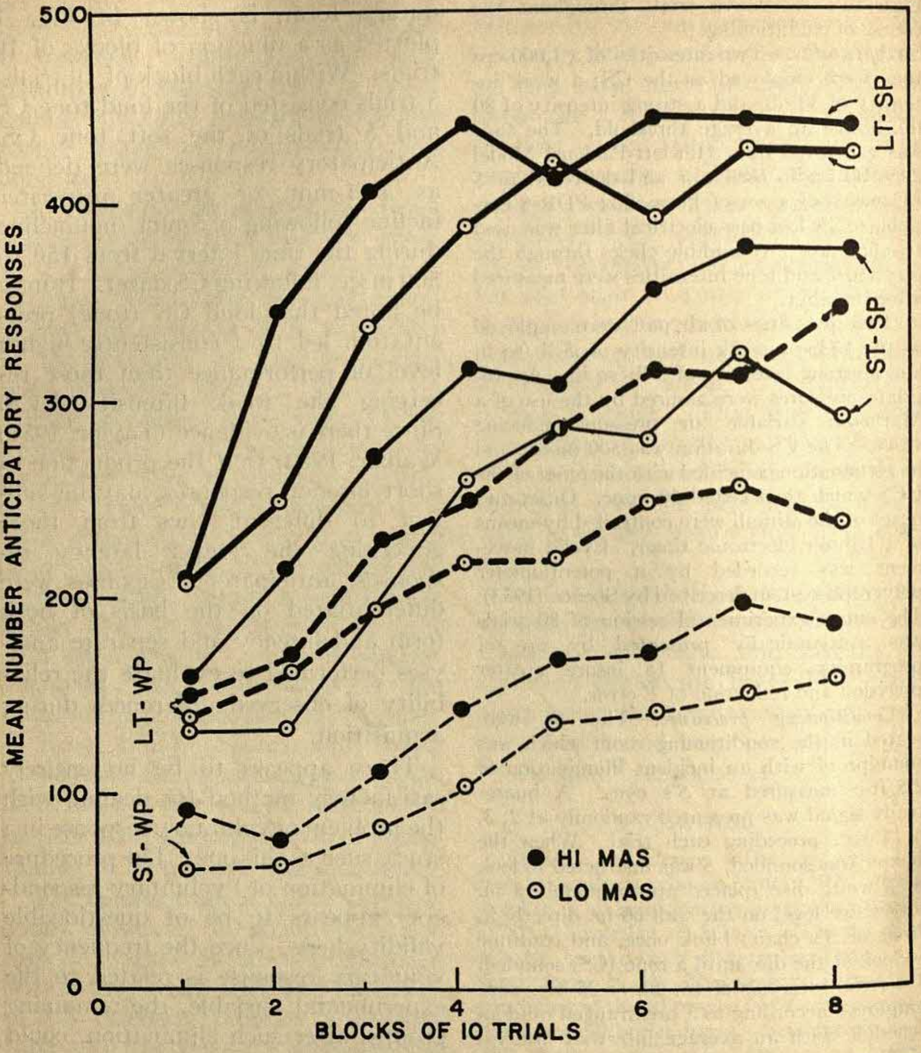


FIG. 1. Mean number of anticipatory responses in the interval 150–500 msec. after CS onset during acquisition. (Each block of 10 trials contained five loud and five soft tone CS presentations. The following abbreviations are used in the figure: ST-WP, soft tone-weak puff; LT-WP, loud tone-weak puff; ST-SP, soft tone-strong puff; LT-SP, loud tone-strong puff.)

appears highly probable that this measure contains less variance attributable to the frequency of voluntary response than either of the other alternatives. Inspection of the data indicates that it introduced no serious distributional problems. Furthermore, in one important sense it has the same meaning for all *S*s: percent CRs on those trials when a CR had the opportunity to occur.[3] Specifically the measure was:

---

[3] These calculations were carried out separately for loud tone presentations (40 trials) and for soft tone presentations (40 trials).

Corrected percent CR

$$= \frac{\mathrm{CR}}{\mathrm{Tt} - \mathrm{tvr}} \times 100$$

where Tt is the total number of trials and tvr is the number of trials on which voluntary responses occurred. For this purpose voluntary responses were defined according to the form criterion (Spence, 1958).

Figure 2 presents this corrected percent CR data for 80 acquisition trials plotted as a function of blocks of 10 trials. As noted before, superior levels of performance occurred to loud tone CS presentations for each of the

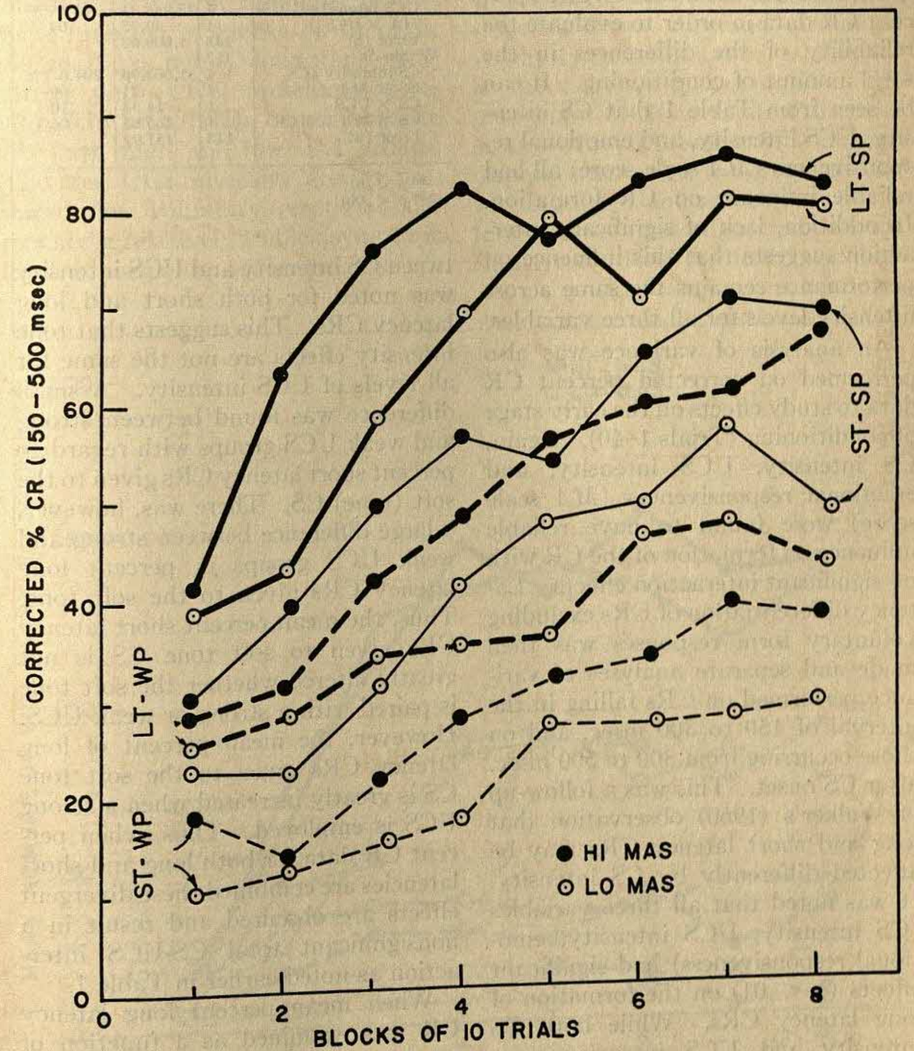


FIG. 2. Corrected percent CR in the interval 150–500 msec. after CS onset during acquisition. (Each block of 10 trials contained five loud and five soft tone CS presentations. The following abbreviations are used in the figure: ST-WP, soft tone-weak puff; LT-WP, loud tone-weak puff; ST-SP, soft tone-strong puff; LT-SP, loud tone-strong puff.)

four groups. The two strong UCS (puff) intensity groups were found to demonstrate superior levels of performance and the highest single level of performance was reached for the High-*MA* group given the more intense CS and UCS condition.

Table 1 summarizes the results of a factorial analysis of variance (Lindquist, 1953) based on the corrected percent CR data in order to evaluate the reliability of the differences in the total amount of conditioning. It can be seen from Table 1 that CS intensity, UCS intensity, and emotional responsiveness (*MA* scale score) all had reliable influence on CR formation. In addition, lack of significant interaction suggests that this influence on performance remains the same across intensity levels for all three variables.

TABLE 1

ANALYSIS OF VARIANCE FOR THE FOUR GROUPS DURING ACQUISITION (Trials 1–80) ON CORRECTED PERCENT CRs OCCURRING IN THE TIME INTERVAL 150 TO 500 MSEC. AFTER CS ONSET

Source	*df*	*MS*	*F*
Between-*S*s	151		
MA scale score (*MA*)	1	8,611.59	8.53**
UCS intensity (UCS)	1	43,968.21	43.55***
MA × UCS	1	42.75	.04
Error (b)	148	1,009.68	
Within-*S*s	152		
CS intensity (CS)	1	31,002.96	204.07***
CS × *MA*	1	.21	.00
CS × UCS	1	14.33	.10
CS × *MA* × UCS	1	465.05	3.06
Error (w)	148	151.92	

** $p < .01$.
*** $p < .001$.

An analysis of variance was also performed on corrected percent CR data to study effects on the early stage of conditioning (Trials 1–40). Again, CS intensity, UCS intensity, and emotional responsiveness (*MA* scale score) were found to have reliable influence on formation of the CR with no significant interaction effects. Latency differentiation of CRs excluding voluntary form responses was then made and separate analyses of variance performed on CRs falling in the interval of 150 to 300 msec. and on those occurring from 300 to 500 msec. after CS onset. This was a follow-up of Walker's (1960) observation that long and short latency CRs may be affected differently by CS intensity. It was noted that all three variables (CS intensity, UCS intensity, emotional responsiveness) had significant effects ($p < .01$) on the formation of long latency CRs. While both CS intensity and UCS intensity were found to affect significantly short latency CRs, emotionality of *S* did not significantly affect short latency CRs. A significant interaction between CS intensity and UCS intensity was noted for both short and long latency CRs. This suggests that tone intensity effects are not the same for all levels of UCS intensity. A small difference was found between strong and weak UCS groups with regard to percent short latency CRs given to the soft (tone) CS. There was, however, a large difference between strong and weak UCS groups in percent long latency CRs given to the soft tone. Thus, the mean percent short latency CRs given to soft tone CS is not greatly altered whether the soft tone is paired with a strong or weak UCS. However, the mean percent of long latency CRs given to the soft tone CS is greatly increased when a strong UCS is employed. Thus, when percent CR data for both long and short latencies are combined these divergent effects are obscured and result in a nonsignificant total CS-UCS interaction as noted earlier in Table 1.

When mean percent long latency CR was examined as a function of tone intensity for all four groups, a very small difference in response given to the loud vs. soft tone was noted for the High *MA*-Strong UCS (puff) group. This can be expected, however, from the near ceiling performance of this group which serves to limit the upper range of possible performance and thereby greatly enhance the possibility of occurrence of the obtained three-way interaction effect between CS intensity, UCS intensity, and emotionality.

Figure 3 presents the mean number of voluntary form responses plotted as a function of blocks of 10 trials. As may be noted, loud tone (CS) strong puff (UCS) presentations resulted in high levels of performance for both high and low *MA* groups. The low UCS intensity groups produced few voluntary responses and reached a relatively stable asymptotic level of response occurrence by approximately the second block of 10 trials.

Due to the relatively small number of voluntary responses observed in this experiment, the large number of tied scores which occurred among them, and the extreme deviation of their distribution from normality, the following statistical procedures were employed for analysis of the voluntary response data. A grand median of voluntary responses was established for the entire sample population. The proportion of *S*s in each of the four groups with the number of voluntary responses exceeding the grand median was tabulated and an arc-sine transformation of these four proportions employed. This resulted in a four

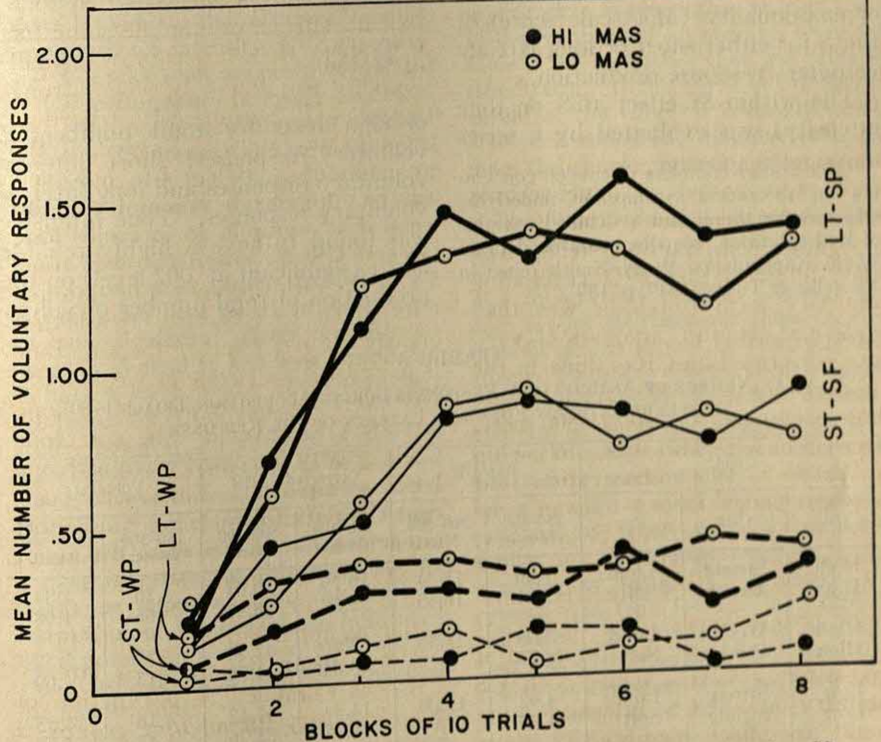


FIG. 3. Mean number of voluntary responses in the interval 150–500 msec. after CS onset during acquisition. (Each block of 10 trials contained five loud and five soft CS presentations. The following abbreviations are used in the figure: ST-WP, soft tone-weak puff; LT-WP, loud tone-weak puff; ST-SP, soft tone-strong puff; LT-SP, loud tone-strong puff.)

cell, single entry table for a simple analysis of variance with use of the variance of the angle as the within-groups estimate of error.[4]

The voluntary form response data were then divided on the basis of latency of occurrence and simple analyses of variance identical to those performed on the total voluntary form data were conducted. The results of these analyses are included in Table 2. It is seen that UCS (puff) intensity reliably influenced the number of both short and long latency voluntary responses produced. Conversely, emotional responsiveness failed to have a reliable effect on the formation of voluntary responses. These analyses gave no indication of any interaction effect between intensity of the UCS employed and level of emotionality (*MA* scale score) of the *S* for either short or long latency voluntary response production.

TABLE 2

ANALYSES OF VARIANCE FOR THE FOUR GROUPS DURING ACQUISITION ON VOLUNTARY RESPONSES OCCURRING IN THE TIME INTERVALS 150–500, 150–300, AND 300–500 MSEC.

Source	*df*	150–500 Msec. *F*	150–300 Msec. *F*	300–500 Msec. *F*
Loud Tone CS				
UCS Intensity (UCS)	1	24.47***	18.87***	16.15***
MA scale score (*MA*)	1	.45	.04	.15
MA × UCS	1	.00	.04	.02
Theoretical within error angle		21.60	21.60	21.60
Soft Tone CS				
UCS	1	17.33***	6.48*	17.35***
MA	1	1.10	.19	1.39
MA × UCS	1	.16	.04	.07
Theoretical within error angle		21.60	21.60	21.60

* $p < .05$.
*** $p < .001$.

TABLE 3

MEAN NUMBER OF ANTICIPATORY RESPONSES DURING ACQUISITION TRIALS 1–80 DIFFERENTIATED BY FORM AND LATENCY OF THE RESPONSE

Cond.			Raw CR		Corrected % CR		Voluntary	
MA	UCS	CS	150–300 Msec.	300–500 Msec.	150–300 Msec.	300–500 Msec.	150–300 Msec.	300–500 Msec.
High	Strong	Loud	7.00	15.34	19.79	50.08	4.26	4.84
High	Strong	Soft	2.21	16.45	5.58	48.97	.66	4.58
High	Weak	Loud	3.08	15.55	7.76	42.34	.79	1.32
High	Weak	Soft	1.55	9.29	3.84	24.34	.13	.60
Low	Strong	Loud	4.92	15.00	14.34	47.42	4.16	4.53
Low	Strong	Soft	1.95	11.97	5.37	35.26	1.45	3.68
Low	Weak	Loud	2.42	11.32	5.10	31.55	1.24	1.55
Low	Weak	Soft	1.21	6.87	3.00	17.84	.32	.79

The within-*S*s effect (CS or tone intensity) was evaluated by a series of sign tests for total number of voluntary responses, short latency voluntary responses and long latency voluntary responses. Tone intensity was found to have a highly reliable effect (significant at .002 level) on the production of total number of voluntary responses and number of short latency voluntary responses for all four groups, and for the two weak UCS (puff) intensity groups in production of long latency voluntary responses. The effects of tone intensity on production of long latency voluntary responses for the strong UCS (puff) groups was in the predicted direction but failed to reach significance.

---

[4] For a more complete discussion concerning the theoretical assumptions underlying this transformation and a detailed example of its application, see discussion of data of W. E. Kappauf and W. N. Smith (cited in Mosteller & Tukey, 1949, p. 189).

Table 3 presents a complete summary of all anticipatory responses differentiated on the basis of both the form and latency of the response. Means based on both the raw and corrected CR data are included for purposes of comparison.

DISCUSSION

From the results of this experiment there is evidence that all three variables employed, CS intensity, UCS intensity, and *MA* scale score are positively related to CR performance in eyelid conditioning. The positive relationship obtained between CS intensity and performance is consistent with the Hullian postulate of stimulus intensity dynamism and has been shown to be of major importance in the determination of CR performance. It was indicated earlier that while CS intensity has been found, in some cases, to have significant effects on performance, in general it has been considered to have a much weaker effect than that produced by UCS intensity. In this study, however, all analyses of the CR data indicate CS intensity to be a strong effect. Nevertheless, it must be noted that the CS (tone) intensity in this design was a within-*S*s effect, unlike previous studies investigating this variable. It might be supposed that the level of adaptation of *S* to the CS was influenced by exposure to both loud and soft CS (tone) stimuli throughout the course of the conditioning.

As Helson (1959) points out, when an *S* experiences several stimuli, the adaptation level (AL) is an integration of both present and residual stimulation. The attainment of this AL also establishes a bipolarity of behavior in such a way that stimuli above the AL tend to elicit one kind of response and those below the AL elicit the opposite type of response. If it is true that the probability of a CR to a stimulus is dependent on the distance of that stimulus from the adaptation level, introduction of two widely dispersed stimuli shifts the AL to some point intermediate between the two. This would further heighten CR responsiveness to the stimulus above adaptation level and further reduce CR responsiveness to the stimulus below adaptation level. This results in greater response differentiation and thereby maximizes performance differences to the stimuli presented. Whether or not adaptation level theory is the most adequate context in which to consider these results, the markedly greater effect in this study as opposed to previous experiments, suggests that at least some kind of contrast effect operates when the same *S* receives more than one level of stimulation. If this turns out to be the nature of the effect, it certainly would involve a somewhat different principle than the simple intensity effect postulated by Hull.

The hypothesis that strong CS and UCS intensity would have a greater effect on performance for High vs. Low *MA* *S*s was not verified. Neither the data for all 80 trials nor those for the first 40, which should be relatively unaffected by ceiling effects, yielded significant interaction at any level. Separate analyses for short and long latency CRs yielded the only evidence of a significant CS-UCS interaction effect. While the mean number of short latency responses produced to the soft tone CS was highly similar for the groups receiving either a strong or weak UCS intensity, the reverse was true for the long latency responses. However, it must be noted that this effect was dependent on a near ceiling performance by the High *MA* group when placed under loud CS and strong UCS intensity conditions.

The results of analyses of the voluntary form data indicate that both CS intensity and UCS intensity are positively related to the production of voluntary responses. Emotionality of *S* was not found to have a reliable influence on the frequency of voluntary responses.

When latency of response was also considered, it was again noted that UCS intensity was positively related to short and long latency voluntary response formation. The finding that CS intensity affects both long and short latency responses is not consistent with the results of both Walker (1960) and Taylor (1954). In both of these studies, CS intensity was reported to influence only short latency voluntary responses and suggested to these authors that short and long latency responses were subject to different laws. The results of the present experiment suggest that voluntary responses falling in either latency interval follow essentially the same pattern of increment throughout the course of conditioning and parallel those of CR formation in a much less pronounced way.

REFERENCES

Baron, M. R., & Conner, J. P. Eyelid conditioned responses with various levels of anxiety. *J. exp. Psychol.*, 1960, **60**, 310–313.

Helson, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. *Sensory, perceptual, and physiological formulations.* New York: McGraw-Hill, 1959.

Hull, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.

King, M. S., Kimble, G. A., Gorman, J., & King, R. A. Replication Report: Two failures to reproduce effects of anxiety on eyelid conditioning. *J. exp. Psychol.*, 1961, **62**, 532–533.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Mosteller, F., & Tukey, J. W. The uses and usefulness of binomial probability paper. *J. Amer. Statist. Ass.*, 1949, **44**, 174–212.

Passey, G. E. The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.

Pavlov, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.

Razran, G. The dominance-contiguity theory of the acquisition of classical conditioning. *Psychol. Bull.*, 1957, **54**, 1–46.

Ross, L. E., & Hunter, J. J. Habit strength parameters in eyelid conditioning as a function of UCS intensity. *Psychol. Rec.*, 1959, **9**, 103–107.

Spence, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Spence, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

Spence, K. W., Haggard, D. F., & Ro L. E. Intrasubject conditioning as a function of the intensity of the unconditioned stimulus. *Science*, 1958, **128**, 774–775. (a)

Spence, K. W., Haggard, D. F., & Ross, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411. (b)

Spence, K. W., & Taylor, J. A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol*, 1953, **45**, 265–272.

Taylor, E. Generalization of the conditioned eyelid response to an auditory stimulus varying in intensity. Unpublished doctoral dissertation, University of Iowa, 1954.

Taylor, J. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

Taylor, J. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Trapold, M. A., & Spence, K. W. Performance changes in eyelid conditioning as related to the motivational and reinforcing properties of the UCS. *J. exp. Psychol.*, 1960, **59**, 209–213.

Walker, E. Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1960, **59**, 303–311.

(Received October 26, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 439–442

AFTEREFFECT OF SLEEP DEPRIVATION[1]

ROBERT T. WILKINSON

Applied Psychology Research Unit, Cambridge, England

24 enlisted men stayed awake one night, slept normally the next, and on the following day carried out either a 1-hr. vigilance task (Experiment I, 12 Ss) or a 30-min. test of serial reaction (Experiment II, 12 Ss). In each case performance fell below control levels (normal sleep on both nights). Ss were their own controls in a balanced design. This "aftereffect" of sleep deprivation was greatest in the morning and, unlike the direct effect of sleep deprivation, was apparent at the start of the test and increased little with time. This suggests disturbed diurnal rhythm as the cause rather than failure to make up for lost sleep.

Almost without exception studies of loss of sleep have been concerned primarily with the "direct effect" of sleep deprivation, that is with the changes which occur during the actual period of sleeplessness. Occasionally, however, measurements have been continued into the periods of recovery following restorative sleep in an attempt to assess what may be called the "aftereffect" of sleep deprivation. These additional measurements are valuable as the only evidence we have so far of this aftereffect, but the conclusions which have emerged are rather contradictory. This is understandable, for such dual-purpose experiments may make it difficult to achieve the necessary experimental control for measuring what may be a relatively small effect. The aim of the present study has been to achieve such control by making the measurement of the aftereffect the main objective. Specifically the question was whether any of the known effects of one night's loss of sleep on visual vigilance (Wilkinson, 1960) and on choice serial reaction (Wilkinson, 1961b) remained after one night of restorative sleep.

METHOD

Tests.—The vigilance test used in Exp. I has been described fully elsewhere (Wilkinson, 1961a). The *S*s watched a screen 6 in. in diameter at a distance of about 2 ft. At an average interval of 3 sec. the screen was faintly illuminated for $\frac{1}{2}$ sec. Occasionally there would be a dim spot of light present in any of 10 possible positions irregularly distributed about the screen. The *S* had to watch for this signal and report its presence by pressing a key. He watched continuously for 1 hr. during which the screen lit 1,200 times. On 32 of these occasions, irregularly spread in time, the signal was present. Continuous "white" noise at 65 db. was used to mask extraneous sound as *S* worked alone in a cubicle.

In Exp. II the test was one of choice serial reaction, Leonard's 5 Choice Test (Leonard, 1959). The *S* is faced with five lights and five associated metal contacts; at the outset one of the lights is on; *S* touches the appropriate contact which puts it out and brings another on; again the associated contact is touched, and so the cycle continues with *S* working at his own speed for 30 min. The order in which the lights come on appears random. Scores are in terms of the number of correct taps, the number of errors (when an incorrect contact is touched, the light still moving on), and finally the number of gaps, i.e., periods of 1½ sec. elapsing between successive responses.

---

[1] The Royal Navy supported this work by providing *S*s, technical research assistants, and financial help through the Royal Naval Personnel Research Committee of the Medical Research Council of the United Kingdom. The work was carried out under the general direction of D. E. Broadbent and with the assistance at various times of R. C. Collecot, P. Grant-Ross, and C. J. Gray, Chief Petty Officers, and R. Edwards, Sick Berth Attendant. All these contributions are most gratefully acknowledged.

Subjects.—Twelve *S*s were used in Exp. I and a further 12 in Exp. II. They were enlisted men between the ages of 18 and 30, and all were volunteers to undergo sleep deprivation.

Sleep deprivation.—This involved staying awake under supervision from 7 A.M. on Monday morning until about 5 P.M. Tuesday, that is around 34 hr. During the sleepless night *S*s amused themselves in their recreation room. The next day was spent in tests of another nature and in routine duties (Exp. I) or in the same test of serial reaction and routine duties (Exp. II). After this *S*s were free to sleep until called at the usual time of 6.30 A.M. the following morning, a period of some 13 hr. No *S* slept all of this time; the usual practice was to sleep for 2 hr. in the late afternoon and then remain awake until about 11 P.M., the normal time for retiring. On the following day, Wednesday, they were tested. This is called the Aftereffect condidition. For the control tests (Normal condition) the program was just the same except that *S*s slept quite normally, that is on the Monday as well as on the Tuesday night.

Design.—In Exp. I *S*s were tested four times on four successive Wednesdays; 6 *S*s were tested in the morning and the other 6 in the afternoon. The *S*s of Exp. II were tested on two Wednesdays (separated by 4 wk.); each *S* was tested both in the morning and the afternoon of each Wednesday. This design and distribution of the conditions for both experiments are shown in Table 1.

RESULTS

Wilcoxon's (1949) nonparametric test was used to calculate the significance levels shown in this section. The influence of the aftereffect of sleep deprivation was seen most clearly in the vigilance test of Exp. I where it reduced the total number of signals reported by about one fifth (Table 2). Out of 12 *S*s 9 showed this effect, 2 showed no change, and only 1 showed a slight opposite tendency, $t(9) = 1$, $p < .01$. The aftereffect of loss of sleep impaired performance at least as much in the first half of the test, $t(10) = 0$, $p < .001$, as in the second, $t(11) = 15.5$, $p < .1$. This is in marked contrast to the direct effect of loss of sleep, which is absent early in the test but increases with time spent at work (Wilkinson, 1960). The aftereffect was greater among the 6 *S*s tested in the morning than in the 6 tested in the afternoon, although this trend was insignificant.

The results in the test of serial reaction used in Exp. II are in the same direction as those in vigilance but not so well defined (Table 3). The aftereffect reduced the total number of corrects and also the total

TABLE 1

EXPERIMENTAL DESIGNS OF THE TWO EXPERIMENTS

Exp. I	Day 1	Day 2	Day 3	Day 4
6 *S*s 6 *S*s	Aftereffect Normal	Normal Aftereffect	Aftereffect Normal	Normal Aftereffect
Exp. II	Day 1		Day 2	
	A.M.	P.M.	A.M.	P.M.
6 *S*s 6 *S*s	Aftereffect Normal	Aftereffect Normal	Normal Aftereffect	Normal Aftereffect

Note.—Aftereffect = normal sleep the night before but no sleep the night before that. Normal = normal sleep on both previous nights but otherwise the same routine.
The days were separated by 1 wk. in Exp. I and 4 wk. in Exp. II.

TABLE 2

Average Signals Seen per *S* in Vigilance

	Normal Sleep	Aftereffect of No Sleep
1st half of the test	10.0	8.0
2nd half of the test	8.3	6.9
Whole test	18.3	14.9

errors. These trends did not reach significance but, so far as they go, they suggest that the aftereffect caused *S*s to trade speed for accuracy. The clearest influence of the aftereffect appeared in gaps, (periods of 1½ sec. during which *S* failed to respond) which is the measure which also shows the greatest direct effect of sleep deprivation. In the morning tests significantly more gaps were recorded under the aftereffect than in the tests under normal sleep, $t(11) = 13$, $p < .05$. In the afternoon tests this effect was reduced to very little. It will be recalled that a similar diurnal trend was observed in Exp. I. As in the first experiment also, the aftereffect was at least as marked in the first half of the test (when 10 out of 12 *S*s showed it during the morning tests) as in the second half.

TABLE 3

Average Scores in the Five-Choice Test of Serial Reaction

	Morning		Afternoon	
	Normal Sleep	After-effect of No Sleep	Normal Sleep	After-effect of No Sleep
Gaps				
1st half	10.2	17.2	13.3	17.2
2nd half	17.8	24.4	25.4	25.0
Total	28.0	41.6	38.7	42.2
Corrects				
1st half	1464	1419	1502	1442
2nd half	1427	1405	1442	1411
Total	2891	2824	2944	2853
Errors				
1st half	31.0	26.2	35.4	26.9
2nd half	33.8	23.5	35.2	29.6
Total	64.8	49.7	70.6	56.5

Discussion

Previous attempts to assess the aftereffect of sleep deprivation have produced contradictory conclusions. Three studies reported an adverse aftereffect on the day following the sleep which terminated the period of deprivation (Edwards, 1941; Smith, 1916; Williams, Granda, Jones, Lubin, & Armington, 1962). Three others reported no effect (Robinson & Hermann, 1922; Robinson & Richardson-Robinson, 1922; Tyler, 1955), and a further three authors remained noncommittal (Armington & Mitnick, 1959; Williams & Lubin, 1959; Williams, Lubin, & Goodnow, 1959). In none of these reports, however, were any significance levels ascribed to the results, and, in general the experimental designs were unsatisfactory. This was due mainly to the assessment of the aftereffect being a secondary objective in experiments whose main aim was to measure the direct effect of loss of sleep. In particular it was difficult to distinguish a true aftereffect from the influence of repeated testing on the same test. This factor can either improve performance through learning, or, as is more likely with tests which are vulnerable to loss of sleep, impair performance due to the declining novelty of the experimental situation (Wilkinson, 1961b). When the same *S*s are repeatedly exposed to a test before, during, and after a period of sleep deprivation such influences make comparisons of pre- and postdeprivation performance tenuous; and even if a separate group is used having the same program of tests but no sleep deprivation this may be unsuitable as a control because of differential practice effects with and without sleep. One solution, that of the present Exp. I, is to omit measurement altogether during the period of deprivation itself. A second (Exp. II) is to use *S*s as their own controls but to counterbalance the order of the experimental and control tests in order to achieve a better balance of the effects of practice. With these precautions the present study has shown in two independent experiments that two tasks known to be particularly sensitive to the direct effects of sleep deprivation are still impaired on the day following restorative sleep. This allows us to say with some confidence that a normal night's sleep does not restore the moderately sleep deprived man to normal.

Still in doubt, however, are the reasons for this. If the body's resources are depleted in some way by prolonged waking activity it may be that additional sleep is essential to restore them to normal by the following day. On the other hand our observed aftereffect could be due to the disruption of diurnal physiological rhythms by sleep deprivation, as Kleitman (1939) found in the case of body temperature. The latter explanation is preferred because of two features of the present results. The first is the suggestion in both experiments that the adverse influence of the aftereffect is less in the afternoon than in the morning. No sleep was taken in the interval and so the best explanation may be that normal physiological rhythms were disturbed in the morning, due perhaps to deeper or more prolonged restorative sleep the night before, and had become partially re-established by the afternoon.

Secondly, it is interesting that the *way* in which performance was impaired by the aftereffect was unlike that of the direct effect of sleep deprivation. Instead of developing gradually with time spent on the test the aftereffect was apparent early in the test and changed little as the test wore on. There is some evidence that changes in performance due to diurnal rhythm per se have this characteristic (Colquhoun, 1962), and this further reinforces the impression that these influences were the main cause of the aftereffect of sleep deprivation we have observed. It may be that when sleep has been lost an aftereffect is best avoided, not by taking additional sleep, but by resuming normal routines as quickly as possible.

REFERENCES

ARMINGTON, J. C., & MITNICK, L. L. Electroencephalogram and sleep deprivation. *J. appl. Physiol.*, 1959, **14**, 247–250.

COLQUHOUN, W. P. Effets d'une faible dose d'alcool et de certains autres facteurs sur la performance dans une tâche de vigilance. *Bull. Cent. Etud. Rech. Psychotech.*, 1962, **11**, 27–44.

EDWARDS, A. S. Effects of the loss of 100 hours of sleep. *Amer. J. Psychol.*, 1941, **54**, 80–91.

KLEITMAN, N. *Sleep and wakefulness.* Chicago: Univer. Chicago Press, 1939.

LEONARD, J. A. 5 choice serial reaction apparatus. *Med. Res. Council Appl. Psychol. Res. Unit Rep., Great Britain,* 1959, No. 326/59.

ROBINSON, E. S., & HERMANN, S. O. Effects of loss of sleep. *J. exp. Psychol.*, 1922, **5**, 19–32.

ROBINSON, E. S., & RICHARDSON-ROBINSON, E. Effects of loss of sleep. II. *J. exp. Psychol.*, 1922, **5**, 93–100.

SMITH, M. A. Contribution to the study of fatigue. *Brit. J. Psychol.*, 1916, **8**, 327–350.

TYLER, D. B. Psychological changes during experimental sleep deprivation. *Dis. ment. nerv. Sys.*, 1955, **16**, 2–8.

WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid, 1949.

WILKINSON, R. T. The effect of lack of sleep on visual watch-keeping. *Quart. J. exp. Psychol.*, 1960, **12**, 36–40.

WILKINSON, R. T. After-effect of loss of sleep on visual watch-keeping. *Med. Res. Council Roy. Naval Personnel Res. Committee Rep., United Kingdom,* 1961, No. OES 366. (a)

WILKINSON, R. T. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. *J. exp. Psychol.*, 1961, **62**, 263–271. (b)

WILLIAMS, H. L., GRANDA, A. M., JONES, R. C., LUBIN, A., & ARMINGTON, J. C. EEG frequency and finger pulse volume as predictors of reaction time during sleep loss. *EEG clin. Neurophysiol.*, 1962, **14**, 64–70.

WILLIAMS, H. L., & LUBIN, A. Impaired performance in a case of prolonged sleep loss. Paper read at the Psychiatric Association Conference, Chicago, 1959.

WILLIAMS, H. L., LUBIN, A., & GOODNOW, J. J. Impaired performance with acute sleep loss. *Psychol. Monogr.*, 1959, **73**(14, Whole No. 484).

(Received November 6, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 443–449

TWO-CHOICE PROBABILITY LEARNING IN THE RAT AS A FUNCTION OF INCENTIVE, PROBABILITY OF REINFORCEMENT, AND TRAINING PROCEDURE[1]

CHARLES N. UHL[2]

University of Alberta

Albino rats were trained in a 2-choice probability learning situation for 1,000 trials. A 4 × 4 × 2 factorial design was used incorporating incentive values of 6, 12, 24, and 48% concentrations of a sucrose solution, probabilities of reinforcement of .60, .70, .80, and .90, and a noncorrection vs. a correction training procedure. Results showed terminal response probability to be: (a) higher under noncorrection than under correction, (b) independent of incentive, (c) an increasing function of probability of reinforcement. Rate of learning was more rapid with higher incentives with the noncorrection procedure but not with correction. All groups exceeded the asymptotic response probability predicted by statistical learning theory. It was concluded that something near 1,000 trials is necessary to accurately assess behavior in probabilistic outcome situations.

Interest in learning in an uncertain outcome situation has increased sharply since the original studies by Brunswik (1939) and Humphreys (1939). This interest has been reflected in the last decade both by many empirical investigations and the emergence of a variety of mathematical models. The statistical learning theory of Estes (1957) perhaps has given the greatest impetus to research on probability learning. Estes' theory predicts that the asymptotic proportion of responses predicting one of two events will equal the probability of occurrence of that event when the probabilities of the two events sum to unity. This predicted behavior is often referred to as probability matching (PM). There have been many empirical verifications of the PM prediction (e.g., Estes & Straughan, 1954). A characteristic both of Estes' model and the various empirical studies of probability learning has been the neglect of a number of parameters known to affect learning and/or performance in more traditional experimental situations. Prominent among these neglected parameters has been that of the incentive value of reinforcement.

Siegel (1959) has presented three models of behavior in two-choice uncertain outcome situations which include a parameter of utility (subjective value of an outcome) which operates to increase the asymptotic probability of response above PM as incentive value increases. The term incentive as used in Hull-Spence theory will be substituted for utility in this paper. PM is a lower bound of asymptotic response probability which Siegel's models predict when the incentive consists only of knowledge that one's prediction was correct. Rate of learning is not affected by incentive in Siegel's models.

---

[1] Supported in part by Grant APA-71 from the National Research Council of Canada, Associated Committee on Applied Experimental Psychology. The writer wishes to express his appreciation for the assistance of John Millman, Anthony F. Nield, and Lorne Yeudall.

[2] Now at the University of Utah.

Edwards (1956) developed a model of two-choice probability learning which predicts that the tendency to exceed PM will increase as the difference between the probability of an event and .5 increases. Edwards (1961) also has argued persuasively that the PM generalization cannot be evaluated properly without a substantially larger number of trials than the 100–300 commonly used.

The objective of the present experiment was to evaluate parametrically the effects of incentive and probability of reinforcement upon response probability over 1,000 training trials with both correction and noncorrection training procedures.

METHOD

Subjects.—The *S*s were 128 experimentally naive male albino rats (Sprague-Dawley) 150–200 gm. in weight upon arrival at the laboratory.

Apparatus.—All training was conducted in a modified Skinner box enclosed in a sound-insulated, ventilated enclosure with a one-way glass exposing the box to view. The box was a 12-in. cube with the top uncovered. It was constructed of $\frac{3}{8}$-in. plywood painted a uniform gray. The two retractable levers and a liquid dipper were mounted in one wall. Each lever was $1\frac{1}{2}$ in. wide, $\frac{1}{2}$ in. thick, and the top of each was 3 in. above the floor, and each protruded $\frac{3}{4}$ in. into the box. The two levers were 6 in. apart on centers. They were discriminable in spatial location only. Approximately 12 gm. force was required to depress a lever fully. The cup of the liquid dipper was equidistant between the levers and $\frac{1}{2}$ in. above the floor. The dipper dispensed .02 ml. of solution. Two 1-in. pilot lights (6-w. bulb) illuminated the interior of the box. One light was mounted 8 in. above the dipper cup and the second was in the same relative location on the opposite wall.

All *E*-controlled events, i.e., the sequence of reinforced response alternatives and the intertrial interval, were operated by an automatic programing device. Responses were recorded automatically on counters. A Hunter klockounter was used to measure the cumulative time the levers were exposed.

Design.—The experiment was a 4 × 4 × 2 factorial design incorporating incentive values of 6, 12, 24, and 48% concentrations of a sucrose solution, probabilities of reinforcement of .60, .70, .80, and .90, and a noncorrection vs. a correction training procedure, respectively. Sucrose solution concentrations were measured by percentage of weight, i.e., a 12% sucrose solution consisted of 12 gm. granulated sugar and 88 cc water. The percentage concentrations were arranged in equal increments on a log scale. Probability of reinforcement (π) refers to the more frequently reinforced response alternative. The probability of reinforcement of the less frequently reinforced alternative was $1 - \pi$. On every trial one or the other of the two response alternatives was associated with available reinforcement.

Procedure

Pretraining.—Upon arrival at the laboratory *S*s were placed in individual cages and kept on an ad libitum food and water schedule for 10 days. On the eleventh day *S*s were placed on a food deprivation schedule consisting of 12 gm. Purina chow every 24 hr. Water was available at all times. This food and water regimen was maintained for the remainder of the experiment. The *S*s were fed 20 min. after a training session. From the eleventh to the seventeenth day *S*s were handled in pairs for 15 min. per day. On the seventeenth and eighteenth days approximately 50 cc of sucrose solution were substituted for *S*'s water supply at the time of feeding to adapt *S* to drinking the sucrose solution. The solution was of 16.9% concentration, the mean log concentration of the four concentrations used during training later. On the eighteenth day *S* was placed individually in the Skinner box for 15 min. with the levers retracted and the dipper magazine empty.

Dipper training was conducted on Days 19–22. Twenty reinforcements of the 16.9% solution were dispensed daily. An effort was made to reinforce *S* for moving back and forth in the vicinity of the retracted levers. Lever training followed on Days 23–27. Twenty trials were run each day. On each trial one lever was exposed while the other was retracted. The sequence of lever exposure was random within the restriction that each lever be exposed 10 trials per day. The exposed lever was retracted and a reinforcement dispensed immediately following its depression. The lever training procedure was designed to minimize the development of position preferences. A 40-trial test of position preference followed on the twenty-eighth day in which both levers were exposed and a press of either lever always was reinforced. The mean preference was .62 with an *SD* of .08.

Training.—Four *S*s were assigned to each of the 32 cells of the Incentive × π × Correction-Noncorrection factorial on the basis of their position preference. Mean preferences were balanced at .50 ± .03 on the left and right within each cell. A higher π was associated with the left lever for two *S*s in each cell and with the right lever for the other two *S*s in each cell.

Probability training consisted of 40 trials per day for 25 consecutive days. Four sequences of 40 reinforced response alternatives were prepared for each of the four π values. Each sequence was determined randomly within the restriction that the appropriate π be maintained exactly over blocks of 20 trials. A random permutation of the four sequences was drawn for each *S* for successive 4-day periods. The twenty-fifth daily session repeated the sequence used on the twenty-first day.

Under the noncorrection procedure both levers were exposed on every trial and both were retracted for 7.5 sec. immediately upon depression of either lever whether a reinforcement was received or not. Under the correction procedure both levers were exposed on every trial but neither was retracted until the "correct" lever was depressed, at which time both levers were retracted for 7.5 sec.

RESULTS

Response probability.—The *S*'s proportion of responses to the more frequently rewarded lever (*P*) was taken as his response probability on each of 26 daily sessions of 40 trials each. Due to a decrease in the within-groups variance as *P* increased, all *P*s were subjected to an arcsine transformation. Transformed *P*s comprised the data treated in all statistical analyses. The data were subjected to analyses of variance of trends (Edwards, 1960) over the entire 26 days of training and over the

TABLE 1

ANALYSIS OF VARIANCE OF TRENDS OF TRANSFORMED RESPONSE PROBABILITIES

Source	All 26 Days		Final 6 Days	
	df	*F*	*df*	*F*
Incentive (I)	3	2.41	3	0.95
Probability of reinforcement (π)	3	55.35***	3	54.38***
Linear	1	164.26***	1	160.45***
Quadratic	1	0.39	1	3.26
Training Procedure (TP)	1	70.24***	1	83.82***
I × π	9	1.22	9	1.11
π × TP	3	3.15*	3	3.22*
I × TP	3	1.26	3	0.27
I × π × TP	9	1.05	9	1.38
Error (a)	96	(1003)[a]	96	(296)[a]
Days (D)	25	208.14***	5	1.40
I × D	75	1.51	15	1.00
π × D	75	5.61***	15	0.80
Linear	3	51.11***	—	—
Quadratic	3	47.45***	—	—
TP × D	25	10.30***	5	0.37
Linear	1	144.46***	—	—
Quadratic	1	96.01***	—	—
I × TP × D	75	1.74	15	0.50
I × π × D	225	1.35	45	1.17
π × TP × D	75	1.76	15	0.77
I × π × TP × D	225	1.35	45	1.07
Error (b)	2400	(38.8)[b]	480	(30)[b]

[a] Pooled between-*S*s error *MS*.
[b] Pooled *S*s × Days error *MS*.
* $p < .05$.
*** $p < .001$.

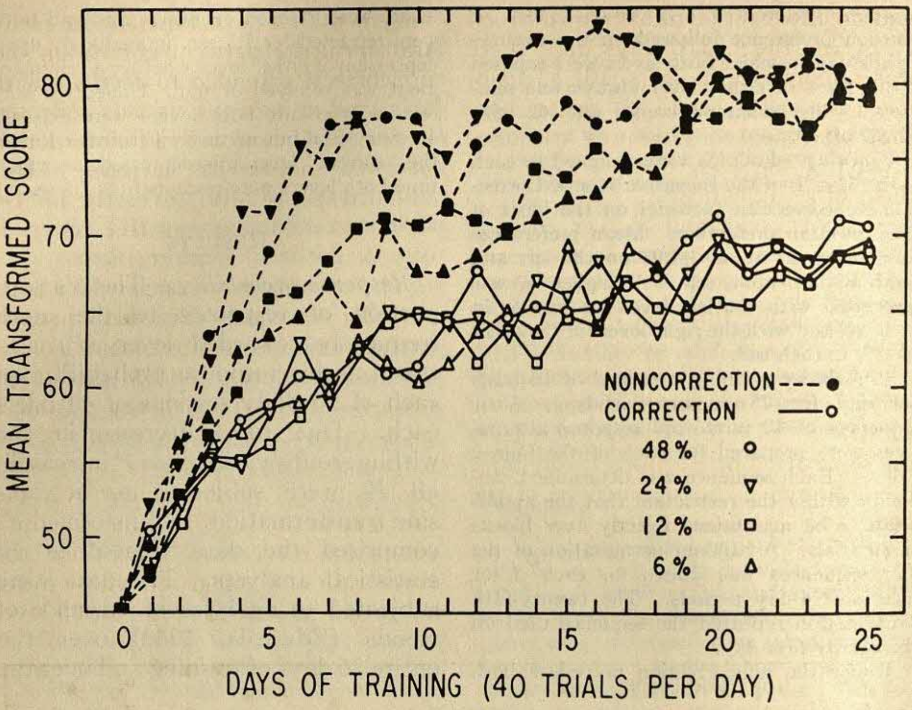


FIG. 1. Response probability as a function of incentive and training procedure.

last 6 days which are summarized in Table 1. (Note that Day 0 was the position preference test day.) Due to the suspected heterogeneity of correlation in the repeated measures over days of training, too many significant *F*s would have been obtained. To offset this bias, no *F* involving days of training was regarded as significant unless it reached a nominal .0001 level.

Figure 1 shows the mean transformed *P* as a function of incentive value of reinforcement and training procedure over 26 days of training. The *S*s trained under the noncorrection procedure showed a more rapid initial increase in transformed *P* and attained a higher terminal level than did *S*s trained under the correction procedure. Statistical support for the difference in rate of learning is found in the TP × D(26) interaction, particularly the quadratic component. The significant main effect of training procedure over the last 6 days reflects the difference in terminal response level. Incentive apparently had no effect upon learning under the correction procedure, but under the noncorrection procedure rate of learning was more rapid with higher incentive values (with the exception that *S*s receiving a 24% solution surpassed those receiving a 48% solution). Statistical evidence for this is found in the significant quadratic component of the I × D(26) interaction for noncorrection *S*s only, $F(3, 1200) = 18.38$, $p < .001$. Although incentive value affected rate of learning under a noncorrection procedure, it had no effect on terminal transformed *P*. Over the last 6 days only there was no significant ($p > .10$) effect of incentive. It should be noted here that all repeated measures variables involving the final 6 days were well short of significance.

Mean transformed *P* over 26 days of training as a function of π and training procedure is presented in Fig. 2. The effect of training procedure has already been mentioned. The general trend is for mean transformed *P* to increase with increasing π. The $\pi \times D(26)$ interaction reflects the separation of π groups throughout the training. The significance of the quadratic component of this interaction was due to a tendency for learning rate to increase as a function of an increase in probability of reinforcement. The difference in mean transformed *P* between correction and noncorrection tended to decrease with an increase in π. This was reflected in the significant $\pi \times TP$ interaction in both analyses of variance. Terminal response level (over the final 6 days) was a linear, positive function of probability of reinforcement.

The terminal response level for both training procedures exceeded π in each probability of reinforcement condition, contradicting the PM prediction of statistical learning theory. Table 2 gives these data and the results of the *t* tests applied. Mean *P* surpassed probability of reinforcement over the last 6 days at the .001 level of significance for all groups except for the correction $\pi = .60$ group whose mean *P* surpassed π at a marginal .05 level

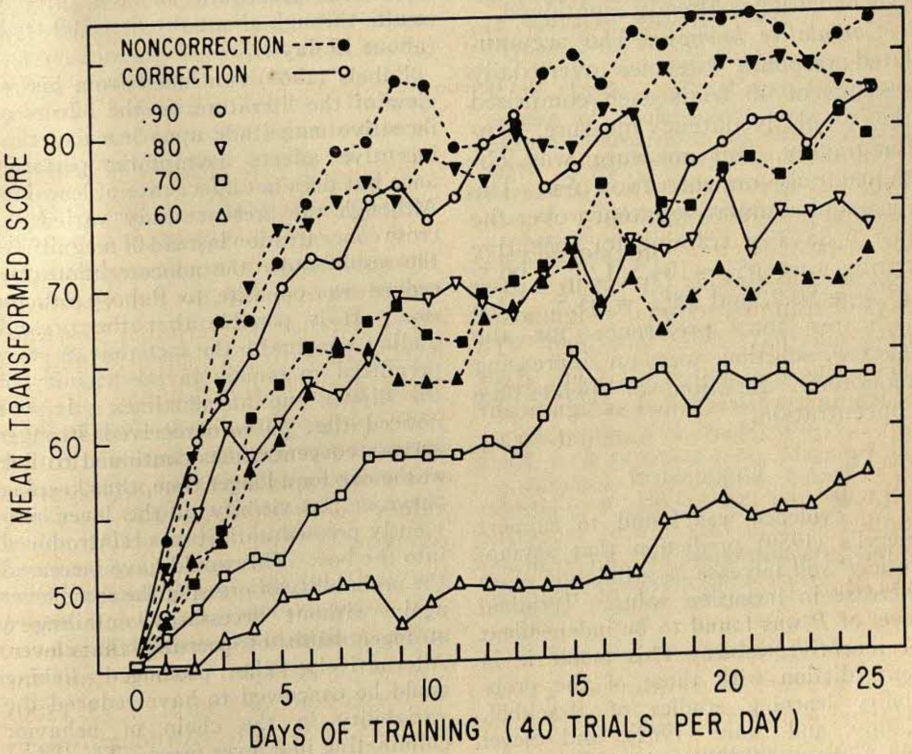


FIG. 2. Response probability as a function of probability of reinforcement and training procedure.

TABLE 2

Mean Differences between Response and Reinforcement Probability ($P - \pi$)

	Probability of Reinforcement			
	.60	.70	.80	.90
Correction	+7.52	+9.75	+12.50	+7.25
t (15)	2.13*	4.13***	5.57***	5.71***
Noncorrection	+28.50	+25.75	+18.00	+8.75
t (15)	7.73***	11.53***	16.04***	25.59***

* $p < .05$.
*** $p < .001$.

of significance. The trend was for these mean differences between P and π to increase from π values of .60 to .80 and to decrease at .90 for correction *S*s, and to decrease with an increase in π for noncorrection *S*s. The difference in these trends as a function of training procedure was reflected in the $\pi \times$ TP interaction in the analysis of variance already mentioned.

Cumulative latency.—The accumulated response latencies over daily sessions of 40 trials each comprised the cumulative latency measure. Unfortunately, this measure was obtained only on correction *S*s. The means of cumulative latency over the last 6 days of training for incentive groups were: 6% = 104.5, 12% = 99.5, 24% = 80.9, and 48% = 65.5. This indicates that preference for the sucrose solution was an increasing monotonic function of percentage concentration.

DISCUSSION

No evidence was found to support Siegel's (1959) prediction that asymptotic P will increase as a function of an increase in incentive value. Terminal level of P was found to be independent of incentive value. This result is in contradiction with those of the probability learning studies of Brackbill, Kappy, and Starr (1962), and Siegel and Andrews (1962), both of which used human *S*s varying the amount of reward. The species difference or variation of amount rather than concentration of reward may account for the discrepancy. However, it seems likely that the much larger number of trials used in the present study may account for the discrepant findings. Incentive value was observed to have an effect upon rate of learning and level of response under the noncorrection procedure in the present results through about the first 700 trials (about 19 days).

Pubols (1960) concluded from his review of the literature on the effects of incentive magnitude upon learning that incentive affects asymptotic performance but does not affect rate of learning. Although the present study varied sucrose concentration instead of magnitude, the result under the noncorrection procedure was opposite to Pubol's conclusion. It is possible that the present finding was due to the fact that *S*s were permitted to remain in the vicinity of the dipper cup at all times. It was noticed that *S*s who received stronger sucrose concentrations continued to lick at the cup for a longer time, thus keeping them in the vicinity of the lever previously pressed until it was reintroduced into the box. This would have increased the probability of pressing the same lever again without necessarily entailing a stronger habit of pressing that lever. Alternatively, the prolonged licking could be conceived to have reduced the variability in the chain of behavior culminating in a lever press. This effect would have been negated under the correction procedure since it was necessary for *S* to move to the other lever if his first response was not reinforced. This interpretation is consistent with the fact that the difference between the correction and noncorrection procedures decreased with an increase in probability of reinforcement. This interaction probably arose through the greater effectiveness of the correction procedure in interfering with positioning behavior under lower probabilities of reinforcement. The fact that *S*s experienced a change in incentive from pretraining to probability training also may have accentuated the incentive effect.

In common with many studies of probability learning with rat *S*s, it was found that terminal response level was higher with a noncorrection procedure than with correction permitted. However, there was some suggestion in the data that the response levels of correction *S*s might have continued to increase if even more trials had been run. The intertrial positioning behavior of *S*s in the present study may account for the finding of an interaction between training procedure and incentive value in rate of learning, unlike the results of other studies in which such positioning behavior was not permitted (Hill, Cotton, & Clayton, 1962).

The PM prediction of statistical learning theory was not upheld even under the correction procedure. The most likely explanation of this result is that PM in rat behavior will not be observed if training trials are continued sufficiently long. The present finding agrees with that of Edwards (1961) who trained human *S*s for 1,000 trials and with that of Wilson (1960) who trained monkey *S*s for 1,024 trials. Edwards' (1956) extreme-asymptote generalization (*P* will exceed probability of reinforcement by greater amounts as the difference between probability of reinforcement and .5 increases) received support. The decline in this function from $\pi = .80$ to $\pi = .90$ was almost certainly due to the ceiling imposed on the $\pi = .90$ group.

REFERENCES

Brackbill, Y., Kappy, M. S., & Starr, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **63**, 32–35.

Brunswik, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.

Edwards, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Rinehart, 1960.

Edwards, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

Edwards, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

Estes, W. K. Theory of learning with constant, variable, or contingent probabilities of reinforcement. *Psychometrika*, 1957, **22**, 113–132.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Hill, W. F., Cotton, J. W., & Clayton, K. N. Effect of reward magnitude, percentage of reinforcement, and training method on acquisition and reversal in a T maze. *J. exp. Psychol.*, 1962, **64**, 81–86.

Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

Pubols, B. H., Jr. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

Siegel, S., & Andrews, J. M. Magnitude of reinforcement and choice behavior in children. *J. exp. Psychol.*, 1962, **63**, 337–341.

Wilson, W. A., Jr. Supplementary report: Two-choice behavior of monkeys. *J. exp. Psychol.*, 1960, **59**, 207–208.

(Received November 16, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 450–453

TRANSFER OF TRAINING FROM A SIMULATED TO A REAL CONTROL SITUATION

M. HAMMERTON

Medical Research Council, Applied Psychology Research Unit, Cambridge, England

This reports an examination of transfer of training when tasks differed only in display appearance; angular control/display relationships were identical. An experimental group of 11 unskilled Ss were trained on a CRT to control the movements of a trolley moving along a miniature railway; a control group of 12 Ss practiced on the real trolley ab initio. Initial transfer was poor (-17% by one measure); the performance decrement being pronounced ($p < .1\%$). Saving, however, was considerable ($>70\%$). Conclusions are: (a) motor response is not immediately transferred, despite identical dynamics and display kinematics, (b) recovery of skill is very rapid, (c) these findings may be attributed to stimulus compounding.

The literature of transfer of training, including that of perceptual-motor skills, is very large (e.g.,Gagné, Foster, & Crowley, 1948; Murdock, 1957). Most of such investigations involving perceptual-motor skills have involved transfer between conditions that did not have identical dynamics. Some workers in the field of simulator training (e.g., Adams, 1957; Dougherty, Houston, & Nicklas, 1957; Newton, 1959) have approached this condition more or less closely, but have usually been interested in skills and situations—such as flying an aircraft—which were so complex that the interpretation of results is difficult.

It is often the object in simulator training to train *S*s to control the movement of an object which, in the real situation, may be at a considerable distance. The simulator designer seeks to reproduce the angular movements of the controlled object, and the dynamics of the control—i.e., the relation between *S*'s controlling movements and the movements of the object. If he succeeds in this, the simulated and real situations will differ in the real size and distance of the controlled object, and in its general appearance and background; but dynamically they will be identical. Under these conditions some workers (e.g., Adams, 1957) would seem to predict almost perfect transfer. However, in a previous study (Hammerton, in press) it had been found that real sizes can be more important in control situations than angular (retinal) sizes. The present experiment was therefore devised to examine transfer between relatively simple and dynamically identical tasks.

The real situation was an acquisition task at a range of 60 yd. The simulated display was presented on a cathode-ray tube (CRT).

METHOD

Subjects.—The experimental group consisted of 11, and the control group of 12 Royal Naval ratings, whose ages ranged from 19–26 yr.

Apparatus and task.—The real object controlled by *S* was a small wheeled trolley, running along a straight length of 3½ in. gauge railway, normal to his line of sight. The *S* and *E* sat in a control cabin 60 yd. from the track. The *S* used a thumb joystick to control the movements of the trolley, the linear velocity of which was proportional, apart from inherent lag, to the deflection of the stick. The control gain was approximately 300 sec.^{-1}, i.e., a stick deflection of 1 cm. produced a velocity of the trolley of 300 cm/sec.

Scoring was by means of a dekatron timer, which *E* started with a press button, and which stopped when the center of the trolley had remained continuously for 2 sec. within a particular zone of the track. The length of this zone was 6 in., and its size and position were indicated to *S* by a white-painted metal target, which appeared behind the track when the dekatron was started and which fell when the dekatron stopped. Fixed to the middle of the trolley was an upright strip of black metal; when this strip was in line with the white target, the trolley was within the target zone. The trolley always started each run from a position 6 yd. to the right of the target. The *S*'s task was to move the trolley to the target as rapidly as possible, and to keep it there until the target fell. His score on each run was the time shown on the dekatron, which included the 2 sec. "holding time."

This real situation was simulated on a CRT, so that, for a given control movement, the angular display movement at *S*'s eye was identical in the simulated and real cases. This was achieved by picking up the trolley's movements directly. The trolley was moved along the track by an endless cable passing over a pulley driven by a velodyne motor. A 70–1 reduction gear was used to connect this pulley to a rotary potentiometer, which controlled the X plane voltage of one output of a double-gun CRT. (To ensure mechanical identity the gear and potentiometer were present during control and transfer, as well as training trials; in the former, however, the potentiometer was on open circuit.) The other output was fixed as a vertical line, and the voltages were so arranged that the "trolley" spot coincided with the fixed line when the trolley was on target. The fixed line thus represented the target: it was made to appear and disappear as the target itself did. By these means the linear movements of the spot upon the CRT were in strict correspondence with those of the trolley itself, and no question of error in representation arose. The distance from starting point to target on the screen was 50 mm. At *S*'s eye this subtended the same angle as the 6-yd. run at 60 yd. range. Thus, whatever control movement *S* made, the angular display movement was identical in both real and simulated situations.

During simulation sessions, the CRT was mounted on a frame outside the control cabin, with its face against the window through which *S* normally looked, the rest of the window being blanked out. The equipment was calibrated twice daily.

Procedure.—The *S*s were shown the apparatus, and the nature of the task and method of scoring were carefully explained to them. The *E* demonstrated the task once, and trials began.

The *S*s in the experimental group carried out 10 trials each day for 3 days using the CRT. They were told the time taken immediately after each trial. They thus had 30 trials, with knowledge of results. On the fourth day, the CRT was removed, and *S*s were given 10 trials in the real situation, again with knowledge of results. The *S*s in the control group were similarly given 10 trials each day in the real situation for 4 days.

RESULTS

The mean performance of the two groups during the course of the experiment is shown in Fig. 1. Each point for the experimental group is thus the mean of 11 *S*s' scores; each of those for the control group is the mean of 12.

It is evident upon inspection that, (*a*) simulator training has produced poor first trial transfer, but (*b*) a considerable saving of training trials in the real situation was effected.

For statistical comparisons of each group on different days, the Wilcoxon signed rank test was used; for comparisons between the two groups the Mann-Whitney *U* test was used (Siegel, 1956). It was found that on the last training trial—i.e., the thirtieth trial—the experimental and control groups do not differ significantly in their performance. The first transfer trial of the experimental group, however, differs very significantly from the last training trial, and from the equivalent (thirty-first) trial of the control group: in both cases $p < .001$. For the experimental group, one measure of transfer used was ϵ, defined as:

$$\epsilon = \frac{F - T}{F - L}$$

where F is the performance on the first training trial, L on the last, and

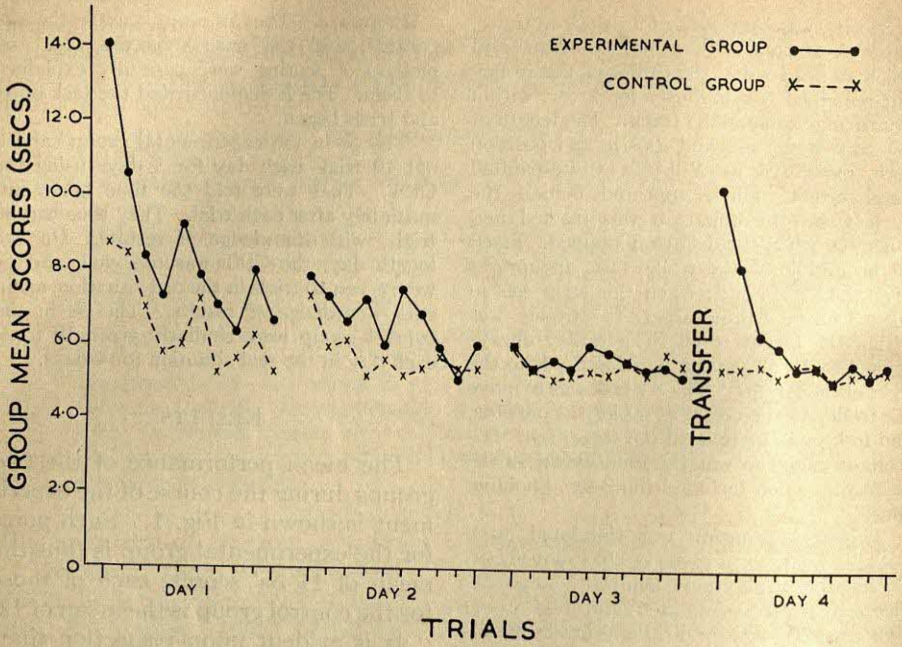


FIG. 1. Transfer from simulated to real task—acquisition times.

T is the performance on the first transfer trial. It was found that individual ϵ values ranged from $-.29$ to $+.91$, with an arithmetic mean of .24. Taking the pooled mean scores—as in Fig. 1.—the overall ϵ was .43. Alternatively, a more usual measure is γ, defined as:

$$\gamma = \frac{C - T}{C}$$

where T is as above, and C is the performance on the first trial of the control group. This gives $\gamma = -.17$, i.e., -17%.

Alternatively, a much cruder measure of transfer, β, would involve the overall mean scores obtained by the control group on Day 1 (A); the overall mean scores obtained by the control group on Day 4 (B); and the overall mean scores obtained by the experimental group on Day 4 (D). Then:

$$\beta = \frac{A - D}{A - B} \times 100$$

The value of β is $+21\%$.

However, there seems to be a considerable saving effect. During the second half of Day 4 the two groups do not differ significantly ($p < .01$). Also, comparing the experimental group for Trials 6–10 on Day 4 with the control group for Trials 6–10 on Day 1, the former are seen to be far more stable. Using the range of each S's scores over these trials as a measure, the two groups differ significantly ($p < .02$, two-tailed). It thus seems reasonable to note that the experimental group's scores did not differ reliably (at the 5% level) from the control group's after three trials on Day 4. The control group's own scores do not differ reliably from one another (again at the 5% level) after their thirteenth trial. This gives a saving of 77% by using simulated training.

Discussion

It is difficult to find accounts of other experiments in which the control dynamics and display kinematics were identical in both training and transfer situations. The striking feature of the present experiment is the poor initial transfer, in spite of that identity. A similar result was reported by Dougherty, Houston, and Nicklas (1957) with their Operational Flight Trainer; but a decrement in performance from a ground trainer to an aircraft can be attributed to a number of factors—e.g., nervousness, proprioceptive changes, etc.—which clearly do not apply in the present case. Here the operator was under no additional pressure, and his situation differed only in the precise nature of the kinematically identical display.

The poor initial transfer, though pronounced, is clearly transient. This is in line with the findings summarized by Bilodeau and Bilodeau (1961), who remark that the half life of a negative transfer effect is "a handfull of trials." However, there are practical situations in which the operator may die if his first-trial transfer is poor, in which case the rest of the learning curve will not be covered.

Learning theory may cast some light on these results. Although the real stimulus was kinematically identical to that in the training system, in many important ways—its detailed appearance and so on—it was perceptually different. Alternatively, we may express this by saying that the real situation includes the original stimulus, but adds irrelevant information. The case is, in effect, one of compounding stimuli, in which performance decrement is to be expected (Estes, 1959). However, having the original response pattern available, *S* can soon fit it to the new situation. In any case, it is clear that something more complex than the straightforward retention of a response pattern is involved.

The practical relevance of the present finding is clearly to cases where "first shot" performance in the real situation is of great importance.

REFERENCES

Adams, J. A. Some considerations in the design and use of dynamic flight simulators. *USAF Personnel Train. Res. Cent. res. Rep.*, 1957, No. 57–51.

Bilodeau, E. A., & Bilodeau, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–280.

Dougherty, D. J., Houston, R. C., & Nicklas, D. R. Transfer of training in flight procedures from selected ground training devices for aircraft. *USN Train. Devices Cent. tech. Rep.*, 1957, No. 71-16-16.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A Study of a Science.* New York: McGraw-Hill, 1959. Pp. 380–491.

Gagné, R. M., Foster, H., & Crowley, M. E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97–130.

Hammerton, M. Control tasks with extended operator-display distances. *Ergonomics*, in press.

Murdock, B. B. Transfer designs and formulas. *Psychol. Bull.*, 1957, **54**, 313–326.

Newton, J. M. Training effectiveness as a function of simulator complexity. *USN Train. Devices Cent. tech. Rep.*, 1959, No. 458-1.

Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

(Received November 19, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 454–460

WORD ASSOCIATION AS A FUNCTION OF CONDITIONED MEANING[1]

HOWARD R. POLLIO[2]

University of Michigan

2 groups of 40 and 37 Ss, respectively, were exposed to the Staats' language conditioning procedure involving 3 nonsense syllables as CSs and 3 sets of 9 meaningful words each, as UCSs. At the conclusion of training Ss gave word associations to the CS syllables. A 3rd group of 25 Ss then rated these associates on the evaluative scale of the semantic differential. Results indicated that SD ratings of the associates produced were congruent with the SD ratings of the paired UCS word group. This was true whether or not the associates were from the original UCS words or whether they were cued by the connotative characteristics of the CS syllable. The results indicate that in the absence of specific word-word associative habits, word association may be determined by the connotative characteristics of the stimulus word.

Although word association is a classical psychological procedure, the critical factors which lead to the evocation of a specific word as an association have yet to be described in any detail. The similarity between word association and memory experiments has been noted by Woodworth (1938), and it is, therefore, not surprising to find that most explanations of word association involve considerations of such factors as frequency and recency which have been shown to have a strong effect on memory for words. Thus, Johnson (1956), Osgood and Anderson (1957), Howes (1957), and Underwood and Schulz (1960) all assume some variety of a frequency of occurrence explanation as sufficient to explain why one word produces another in the word association procedure.

Staats (1961; Staats & Staats, 1959), in his elaboration of the Osgood (Osgood, 1953; Osgood, Suci, & Tannenbaum, 1957) and Mowrer (1954) mediational hypothesis of word meaning acquisition, has assumed that while the frequency of occurrence of an individual word, or its frequency of co-occurrence with another word, provide some basis for predicting overt verbal associative behavior, the conditioned meaning response aroused by stimulus words must also strongly affect the course of overt association. As evidence for this hypothesis, Staats has examined the correlation between the evaluative semantic differential (SD) rating of a stimulus word and the mean evaluative SD rating of the first 20 associates to that stimulus word. For 10 words, the rank-order correlation between these ratings was .90 ($p < .01$).

A more extensive analysis of the semantic relationship between a stimulus word and its first associates has been done (Pollio, 1962) and the results show a significant positive correlation between the SD rating of 50 Kent-Rosanoff stimulus words and the SD ratings of their primary associates on the evaluative, potency, and activity dimensions of the semantic differential. Meaningful mediation would seem to be involved in determining overt word associations although its specific contribution is not completely clear.

---

[1] The present research was done while the author was on a Predoctoral United States Public Health Service Research Fellowship. Arthur W. Melton offered many helpful suggestions during the course of this investigation.

[2] Now at the University of Tennessee.

The purpose of the present experiment is to examine word associations given in response to an originally neutral language stimulus such as a nonsense syllable of low association value which has had a specific meaning response conditioned to it. The procedure successfully used by Staats (Staats, 1961; Staats & Staats, 1959) permits the conditioning of specific meaning responses to nonsense syllables without the concomitant conditioning of any or many words as direct word associates. We have assumed that after language conditioning takes place: (*a*) The *S*s will be able to provide a significantly greater number of meaningful word associations to nonsense syllables which initially aroused few if any associates. (*b*) The SD ratings of associates given in response to a nonsense syllable having meaning responses established through this procedure, will be similar to the mean SD ratings of the words used as the UCS words.

METHOD

Subjects

There were two experimental groups; Group I had 40 *S*s while Group II had 37 *S*s. A third group of 25 *S*s was used to rate material obtained from Groups I and II. All *S*s were female introductory psychology students at the University of Michigan and served in this experiment as part of their course requirement.

Materials

The materials consisted of the following three trigrams, with their respective association norms as restandardized by Archer (1960); GEJ (9%), QIH (6%), and XAD (7%). These words will be referred to as CS words. Three sets of nine words each were used as UCS words and had the following characteristics: There was one set of nine UCS words with positive evaluative SD ratings as determined by the Jenkins, Russell, and Suci (1958) norms, one set of nine with negative evaluative SD ratings, and one set of nine with neutral evaluative SD ratings.

The nine words serving as the positive UCS words and the nine words serving as the negative UCS words were so selected as to have fewer within-set common associates than between-sets common associates. There are 36 intragroup associates common to two or more words in the set of positive UCS words (i.e., $M = 4.00$); while there are 42 intragroup associates in the set of negative UCS words ($M = 4.07$). The communality between these two sets yields 87 cases ($M = 4.85$). This control on the interword associative tendencies of the UCS words serves to guarantee that if the SD profiles of associations emitted by *S*s in response to one of the nonsense syllables are similar to the SD profiles of the UCS words paired with that nonsense syllable, this similarity is the result of communality in meaning response, rather than the result of communality in intralist associative relations such as have been described by Deese (1959) and by Rothkopf and Coke (1961).

The specific words employed in the various sets of UCS words will be found in Table 1. The 1958 SD ratings were taken from the Jenkins, Russell, and Suci (1958) norms while the second set of SD ratings presented was obtained from the 25 *S*s in our word rating group. As can be seen from Table 1 the present ratings, when compared to those obtained by Jenkins et al., produce both less polarized mean SD values and larger within-group sigmas for all three sets of UCS words. The differences in within-group variability are particularly noticeable for neutral UCS words. While the mean difference between ratings is not great, many of the individual ratings in this group deviate considerably from the original norms (e.g., GLOW).

Procedure

Meaning conditioning.—Two different conditioning sessions were run, each with a different group of *S*s. For Group I, GEJ was paired with positive UCS words; XAD was paired with negative UCS words; and QIH was paired with neutral UCS words. For Group II, XAD was paired with positive UCS words, GEJ with negative UCS words, and QIH was again paired with neutral UCS words. The specific pairing of CS syllables and UCS words took place as follows: A CS syllable was flashed on a screen for 5 sec.

TABLE 1

UCS WORDS AND THEIR EVALUATIVE SEMANTIC DIFFERENTIAL RATINGS

Positive Words			Neutral Words			Negative Words		
UCS Word	1958 Rating[a]	Present Rating	UCS Word	1958 Rating[a]	Present Rating	UCS Word	1958 Rating[a]	Present Rating
MUSIC	1.60	1.60	ARCHED	4.07	3.70	SICKNESS	6.50	6.55
EATING	1.57	2.75	BOULDER	4.17	4.00	COLD	5.27	4.55
HOUSE	1.77	2.60	HEAVY	4.10	4.80	ROUGH	5.00	4.80
COMFORT	2.03	2.05	BRISTLY	3.93	4.75	ANGER	5.57	4.63
SMOOTH	2.57	2.60	GLOW	3.93	1.70	SOUR	5.37	5.30
SWEET	1.93	1.90	LONG	3.87	4.45	TROUBLE	6.10	5.45
BEAUTIFUL	2.00	1.80	RUN	3.87	3.55	HUNGRY	5.17	5.15
HEALTH	1.97	1.25	MALLET	3.80	4.75	THIEF	6.37	5.75
SWIFT	2.73	2.25	BLOCK	3.67	4.40	AFRAID	5.47	5.50
M	2.00	2.09		3.93	4.01		5.62	5.30
SD	.35	.48		.14	.93		.49	.59

[a] Taken from the Jenkins, Russell, and Suci (1958) norms.

After 1 sec., *E* read one of the UCS words aloud, and all *S*s pronounced the word after him. This procedure was followed until each of the three CS syllables had been paired with nine appropriate UCS syllables. This made a total of 27 pairings in all. The order of presentation of the UCS-nonsense-syllable pairings was random for both groups. During this conditioning period *S*s were told not "to make any special attempt to memorize either the words or the syllables."

Word association.—At the conclusion of the training sessions, *S*s in both groups were asked "to write down, in the booklets given to you as you entered, the first word that comes into your mind when you see the nonsense syllable." These booklets contained 6 pages. On each of these pages the three CS syllables appeared in one of the six permutations possible for three elements. For every *S*, each of the six possible orders was presented; the specific order of the permutations being randomized across *S*s and groups. Association was paced, so that *S*s were given approximately 20 sec. in which to respond to any set of three syllables. Only those words given in the first 20 sec. association are considered in subsequent analyses. These words are most appropriate for demonstrating the effects of conditioned meaning on word association uncontaminated by the nature of the specific word associates already produced. Choosing the associates produced in any other association period, or the mean value for all six periods, adds in the confounding effect of these preceding associates. Since the purpose of this experiment is to examine the effects of conditioned meaning and not of prior associative responses, on overt word association only those associates produced in the first 20 sec. association period provide appropriate data for this analysis.

Independent rating of associates.—All of the word associations given by the 77 *S*s of

TABLE 2

MEANS AND *SD*S OF THE SD RATINGS OF THE ASSOCIATIONS GIVEN IN RESPONSE TO CS SYLLABLES AS A FUNCTION OF CONDITIONED MEANING PAIRING

Experimental Group	Paired UCS Word Group						*N*
	Positive		Neutral		Negative		
	M	*SD*	*M*	*SD*	*M*	*SD*	
I	2.76	1.00	3.10	1.13	4.72	1.27	40
II	2.79	1.48	3.67	1.25	4.39	1.37	37

Groups I and II during this first 20-sec. period were given to a different group of 25 female *S*s for rating of each word on the good-bad evaluative scale of the semantic differential. The specific order of words to be rated was randomly determined in booklets presented to this group of *S*s. In each of these booklets, the associates produced by *S*s in Groups I and II were typed in capital letters. A 7-step evaluative SD scale, good-bad, was printed under each of these words. The *S*s indicated their ratings of a word by placing a mark on the scale where they thought appropriate. The instructions given to this rating group were essentially similar to those used by Jenkins, Russell, and Suci (1958).

RESULTS

Evaluative characteristics of all associates.—After SD ratings were secured for the word associates given in response to the CS syllables for Groups I and II, mean SD values were obtained for each of these associations. Table 2 shows that for both groups, the SD rating of the word associates given in response to the CS syllables were similar to the SD ratings of the paired UCS words. Thus, CS syllables paired with positive UCS words produced word associates which had positive evaluative meaning, while CS syllables paired with negative UCS words produced word associates that had negative evaluative meaning. The SD ratings of associates to the neutral CS syllable fell between the ratings of the positive and negative groups.

TABLE 3

ANALYSIS OF VARIANCE OF THE DIFFERENCES IN SD RATINGS OF ASSOCIATIONS AS A FUNCTION OF CONDITIONED MEANING

Source	*df*	*MS*	*F*
Between *S*s	76		
Groups (G)	1	.47	
Error (b)	75	1.35	
Within *S*s	154		
Conditioned Meanings (M)	2	63.61	36.35***
M × G	2	3.96	2.26
Error (w)	150	1.75	
	230		

*** $p < .001$.

An analysis of variance was performed on these data, and the only significant source of variation among means was attributable to conditioned meanings. Thus a CS syllable paired with meaningful UCS words tends to produce associates having the same SD ratings as the UCS words themselves.

Evaluative characteristics of cued associates.—At this point it may be argued that the SD ratings of the associates to the CS syllables were similar to the SD ratings of the UCS words because the associates by and large, were the UCS words themselves. An examination of the associates produced, however, revealed that for the 40 *S*s in Group I, 54% (64 of a possible 120) of these associates were words not contained in any of the UCS word groups, while the comparable figure for the 37 *S*s in Group

TABLE 4

NUMBER AND NATURE OF CUED ASSOCIATES AND MEANS AND *SD*s OF THE SD RATINGS OF THESE CUED ASSOCIATES AS A FUNCTION OF PAIRED UCS WORDS

Experimental Group	Paired UCS Word Group								
	Positive			Neutral			Negative		
	M	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
I	3.10	.93	22	3.28	1.14	20	4.70	1.50	22
II	3.25	1.85	15	3.54	1.17	20	4.00	1.28	18

II was 48% (53 of a possible 111). Thus, about half of all of the associates given could not be attributed to *S* learning a specific UCS word as a paired associate to a CS syllable. A non-UCS word, which was given as an associate to a CS syllable, will be called a cued associate in subsequent discussion. A word was considered as a cued associate if it was not spelled exactly the same as any one of the UCS words. For example, ARCHERY, despite some formal similarity to the UCS word ARCHED, was considered a cued associate.

Table 4 presents the number and nature of the cued associates given as a function of the paired UCS word group, as well as the means and *SD*s of the SD ratings for these cued associates. As can be seen, the means are ordered as expected on the basis of the experimental hypothesis, i.e., cued word associations given in response to the CS syllable paired with positive UCS words tend to be positive, those paired with the negative UCS words tend to be negative, and those paired with neutral UCS words tend to fall between these extreme groups.

TABLE 5

ANALYSIS OF VARIANCE OF THE DIFFERENCES AMONG THE MEAN SD RATINGS OF CUED ASSOCIATES

Source	*df*	*MS*	*F*
Meanings (M)	2	16.460	9.058***
Groups (G)	1	.250	—
M × G	2	2.735	1.505
Error	111	1.817	
Total	116		

*** $p < .001$.

Since the means presented in Table 4 are dependent on each *S*'s responses and since the number of responses per condition are confounded over *S*s, there is no completely adequate statistic to test the significance of the differences among these mean SD ratings. Consequently, a simple 3 × 2 factorial analysis of variance, which in effect disregards within-*S* correlation, was performed on these data. The results of these analyses indicate that the evaluative SD rating of a cued associate to an originally neutral CS nonsense syllable is a function of the evaluative SD rating of the UCS words paired with that CS syllable.

Evaluative characteristics of recalled associates.—Approximately 50% of the associates given in response to

TABLE 6

NUMBER AND NATURE OF THE SPECIFIC UCS WORDS PRODUCED AS ASSOCIATES AS A FUNCTION OF THE CONDITIONED MEANING CHARACTERISTICS OF THE CS SYLLABLE

Experimental Group	Conditioned Meaning Characteristics of CS Syllables	Number and Nature of Specific UCS Words Produced as Associates to CS Syllables			Total
		Positive	Neutral	Negative	
I	Positive	17	0	1	18
	Neutral	13	6	1	20
	Negative	1	5	12	18
II	Positive[a]	18	2	1	21
	Neutral[a]	5	6	5	16
	Negative[a]	2	2	14	18

[a] A different *S* in each of these conditions failed to produce an associate in the 20-sec. association period.

each of the three nonsense syllables for both groups were the UCS words themselves. Table 6 presents the semantic characteristics of the specific UCS word recalled as a function of the connotative characteristics of the CS syllables. With the sole exception of the UCS words recalled in response to the CS syllables conditioned neutral, the UCS words recalled had the same connotation as was conditioned to the CS syllable. This effect is quite strong for both the positively and negatively conditioned syllables. A chi square test, done on the 3 × 3 table obtained by collapsing across experimental groups, yielded a value of 63.46 ($df = 4$; $p < .001$). These results suggest that the semantic characteristics of the UCS word recalled as a paired associate to a CS syllable tends to have the same connotation as was conditioned to the CS syllable. Discrimination of the evaluative characteristics of the UCS words seems to have occurred, such that intrusions from the opposite evaluative UCS word class do not occur.

Discussion

The results of the present study indicate that the meaning response aroused by a verbal stimulus is an important factor affecting the evocation of associative responses. We have seen that the evaluative meaning of a word associate is similar to that of its evoking stimulus. This implies, as both Osgood (1953) and Staats (1961) have suggested, that meaning responses often precede, and to a large extent, determine the connotative characteristics of associated responses.

Contrary to Staats' supposition that the language conditioning procedure provides a technique for the conditioning of meaning without the concomitant conditioning of specific words to CS syllables, the present results indicate that about 50% of the UCS words are recalled by *S*s. Using differences in the SD ratings of CS syllables after language conditioning as a dependent variable, is, therefore, open to the criticism that it is the UCS words which are recalled, and are being rated, rather than the CS syllable itself. It is, of course, possible to argue that UCS syllables are recalled *because* of the meaning response conditioned to the CS syllable, i.e., if associations are determined by the meaning response aroused by the CS syllable then the UCS words are produced because they most nearly approximate the meaning response now conditioned to the CS.

The results presented in Table 6 offer some support for this hypothesis by showing that there are few intrusions in recall from UCS word groups having different evaluative characteristics than those conditioned to the CS syllable. Where specific UCS syllables are recalled, they are congruent in connotation to the evoking CS syllable. For syllables conditioned neutral, however, intrusions from other UCS lists did occur. The single UCS word most often inappropriately intruded as an associate to the neutral CS syllable was HOUSE. An examination of many of the words contained in the neutral UCS word group (ARCHED, BOULDER, HEAVY, MALLET, BLOCK) indicates that semantic attributes other than evaluative (e.g., potency) may have been conditioned to this syllable, and that the recall or association of the word HOUSE may have been made in response to these attributes. The recall of UCS words given in response to the neutral CS syllable then follows the same rule as does recall given in response to the positive and negative CS syllables.

Although this analysis provides a possible explanation for the pattern of recall of UCS words in the present study, it does not answer the methodological criticism raised earlier. At present, the operations involved in the language conditioning procedure do not permit us to determine if changes in the SD ratings of CS syllables are the result of conditioned meaning responses or are the result of simple paired-associate recall of UCS words.

The results involving cued word

associates, however, clearly demonstrate the effect of conditioned meaning responses on word association. Here, associative responses were not contained in any of the original UCS word lists, although they were similar in evaluative connotation to these words. It seems that when *S*s are unable to utilize specific word-word connections (as occurs in in paired-associate recall) they produce word associations largely on the basis of conditioned meaning. Perhaps the absence of direct word-word connections is a necessary condition for the utilization of conditioned meaning as the primary determiner of overt word association.

These analyses suggest that both recall and association are affected by the meaning response common to a group of words paired with an originally neutral stimulus. In terms of recall, the present results indicate that intrusion errors occur only infrequently if the to-be-recalled material is homogeneous with respect to a shared connotative dimension. In terms of word association, the present results indicate that in the absence of strong word-word connections among words, stimulus-aroused meaning responses precede and determine subsequent associated responses. Under this condition associative responses are largely the result of similarity in meaning between the stimulus and response words.

REFERENCES

ARCHER, J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22.

HOWES, D. On the relationship between the probability of a word as an association and its general linguistic usage. *J. abnorm. soc. Psychol.*, 1957, **54**, 75–85.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

JOHNSON, D. Word association and word frequency. *Amer. J. Psychol.*, 1956, **69**, 125–126.

MOWRER, O. H. The psychologist looks at language. *Amer. Psychologist*, 1954, **9**, 660–692.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., & ANDERSON, L. Certain relations among experienced contingencies, associative structure, and contingencies in encoded messages. *Amer. J. Psychol.*, 1957, **70**, 411–420.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

POLLIO, H. R. Word association as a function of semantic structure. Unpublished doctoral dissertation, University of Michigan, 1962.

ROTHKOPF, E., & COKE, E. The prediction of free recall from word association measures. *J. exp.Psychol.*, 1961, **62**, 433–438.

STAATS, A. W. Verbal habit families, concepts, and the operant conditioning of word classes. *Psychol. Rev.*, 1961, **68**, 190–204.

STAATS, A. W., & STAATS, C. K. Meaning and *m*: Correlated but separate. *Psychol. Rev.*, 1959, **66**, 136–144.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received November 23, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 461–465

EFFECT OF GOAL-BOX SIMILARITY ON THE AFTEREFFECT OF NONREINFORCEMENT AND RESISTANCE TO EXTINCTION[1]

E. J. CAPALDI AND JAMES E. SPIVEY

University of Texas

4 groups of rats were trained to traverse a straight alley, 3 under the same pattern of partial reinforcement, 1 under consistent reinforcement. The partial groups were given intertrial reinforcement (ITR) following nonreinforced trials preceding reinforced trials. ITR consists in placing S by hand in the baited goal box during the intertrial interval. It was found, as hypothesized, that as similarity between the training goal box and the ITR goal box decreased, resistance to extinction increased. It was concluded that the availability of the nonreinforced aftereffect for conditioning on some trial depends upon the nature of the activities occurring between trials and that the proprioceptive consequences of the instrumental response function as aftereffects.

The reinforced (R) trial of an instrumental learning situation serves in at least two capacities: it replaces a previous aftereffect of nonreinforcement (S^N) occasioned by a nonreinforced (N) trial with the aftereffect of reinforcement (S^R), and it increases the capacity of whatever stimuli are present to evoke the instrumental response (Hull, 1952; Sheffield, 1949). The replacement but not, of course, the reinforcement capacity is served by the N trial. Capaldi, Hart, and Stanley (1963) attempted to separate the two capacities of the R trial by means of a procedure termed intertrial reinforcement (ITR). The ITR procedure consists in placing *S* by hand in the baited goal box during the intertrial interval. It was reasoned that while an ITR trial could serve to replace S^N with S^R, it could not serve as a reinforcer, at least with reference to the instrumental response, because, of course, the response had not been executed.

According to the aftereffects hypothesis, S^N is conditioned to the instrumental response on R trials which follow N trials. Obviously, partial reinforcement acquisition does, and consistent reinforcement acquisition does not, allow S^N to be conditioned to the instrumental response. In extinction, only S^N occurs. It follows that partial reinforcement will be followed by greater resistance to extinction than will consistent reinforcement. It also follows that if S^N is not conditioned to the instrumental response, partial reinforcement will not be followed by greater resistance to extinction than consistent reinforcement.

In the Capaldi et al. (1963) investigation, ITR was administered to a consistent group and to two partial groups in the training goal box, i.e., the goal box *S* entered from the runway. The partial group which received ITR following N trials which preceded R trials was less resistant to extinction than the partial group which received ITR following R trials and no more resistant to extinction than the consistent group. From what was said above, it should be clear that these results support the aftereffects hypothesis and the assumption that ITR replaces S^N with S^R. In the present experiment, all partial groups were given ITR following N trials preceding R trials. The groups differed in that the ITR goal box was more or less similar to the training goal box. The hypothesis under investigation was that the capacity of S^R to replace S^N would decrease as the similarity between goal boxes decreased.

---

[1] Portions of this paper were presented at the Psychonomic Society, St. Louis, September 1962.

METHOD

Subjects.—The *S*s, which ranged in age from 89 to 106 days at the beginning of the experiment, were 15 male and 29 female albino rats from the colony maintained by the Psychology Department, the University of Texas. The *S*s were divided into four groups of 11 *S*s each by means of a random procedure.

Apparatus.—The apparatus consisted of a straight-alley runway and a battery of four goal boxes. The straight alley was painted a midgray. Two of the goal boxes were painted black, one white, and one dark gray. The 4-in. wide runway was 7½ ft. long and enclosed by 8-in. high sides, covered by a clear plastic in which small holes were drilled to provide ventilation. An 8-in. long floor treadle suspended over a microswitch was located at the beginning of the alley. Placing *S* on the treadle started a Standard Electric .01-sec. timer. The clock was stopped when *S* broke a photobeam located 4½ in. from the distal end of the alley. A Fiberglas guillotine door located 12 in. from the distal end of the alley was lowered behind *S* to prevent retracing.

The final 4 in. of the right runway wall were removed to provide an entrance door to the goal box which was positioned in front of it. The goal box lay at a 90° angle to the runway. A metal guillotine door at the entrance door could be lowered to prevent retracing. Each of the L shaped goal boxes was 4 in. wide and 8 in. high, with an overall length of 9 in. Bending off to the right at a 90° angle at the end of the goal box was a small, 2 in. wide × 4 in. long section which contained the food cups. The food cup was painted to match the color of the goal box which contained it. One of the black goal boxes was employed on R trials, the other on N trials. The goal boxes were judged from the Munsell neutral value scale to have the following reflectances: white, 90.0%; gray, 6.6%; and black, 3.1%. An unshielded 200-w. bulb was located 6 ft. directly above the goal box positioned at the entrance door. The goal box positioned at the entrance door was covered with a hardware cloth lid hinged to the alley.

Preliminary training.—The *S*s were fed a wet mash for 1 hr. each day in the home cage throughout the experiment. On Days 1 through 10 *S*s were housed in group cages and on Day 11 each *S* was placed in an individual cage. On Days 4 through 7 *S*s were handled outside the home cage in groups of 12 each for 30 min. A wet mash was available during this phase of handling to insure that *S* would eat readily outside the home cage. On Days 8 and 9 *S*s were handled for a full hour and on Day 10 *S*s were not handled.

Experimental training.—Acquisition training was given on Days 11 through 15, six trials per day. On Day 11 all trials ended in reinforcement. On Days 12 through 15 three of the groups were given 50% partial reinforcement, while the fourth remained on consistent reinforcement. The R trials for the partial groups occurred on Trials 1, 3, and 6 (Day 12), Trials 2, 3, and 6 (Day 13), Trials 1, 4, and 6 (Day 14), and Trials 3, 4, and 6 (Day 15). On runway trials *S* always entered the black goal box. The ITR trials occurred for the partial groups following N trials preceding R trials, e.g., following Trials 2 and 5 on Day 12. The consistent group received ITR following the same trials each day as the partial groups. ITR trials were not given on Day 11 of experimental training, the primary purpose of the Day 11 trials being to insure that all *S*s were eating in the goal box prior to the introduction of partial reinforcement and, of course, ITR trials. For one partial group (PB), all ITR trials were given in a black goal box, for another (PG), in the gray goal box, and for another (PW), in the white goal box. The consistent group (C) was given its ITR trials in a black goal box. Observation of *S*s in the goal box indicated that they ate on all R and ITR trials.

A trial was begun by placing *S*'s forepaws on the tip of the floor treadle. If *S* had not broken the photobeam within 60 sec., it was gently pushed down the runway, a time score of 60 sec. being assigned for that trial. The period of goal-box confinement was 15 sec. on all trials. On R trials *S* found a wet mash and on N trials, an empty food cup. The intertrial interval, which was spent in a wire-mesh cage, was 30 sec. On ITR trials *S* spent only the initial 15 sec. of the intertrial interval in the waiting cage, the remaining 15 sec. being spent in the appropriate goal box (black, gray, or white) where its nose was placed over the baited food cup by *E*. Following the confinement period *S* was immediately brought forward to the start of the runway and the next trial was begun.

All *S*s received 18 extinction trials at 6 trials per day in a black goal box. The duration of goal-box confinement, intertrial interval, etc. was as above.

RESULTS

A test for differences between the mean of the log times on the first and the last day of acquisition was made employing analysis of variance. Since in both cases $F < 1$, it can be assumed that the groups did not differ in acquisition either at the beginning or at the end of training.

The mean of the log running times for each group in extinction is presented in blocks of two trials in Fig. 1. Considering all trials, Group C showed the least and Group PW the most resistance to extinction. Group PG showed slightly greater resistance to extinction than Group PB. Only on the initial day of extinction did Group PB differ greatly from Group C. The mean of the log times on Day 1 was tested for differences using analysis of variance. Differences were significant at well beyond the .01 level, F (3, 40) = 9.25. Duncan's multiple-range test (Edwards, 1960) indicated that while none of the partial groups differed from each other on Day 1, each differed from Group C at beyond the .01 level.

An analysis of variance employing the mean of the log times on all extinction trials yielded F (3, 40) = 6.53 which is significant well beyond the .01 level. Duncan's multiple-range test indicated that differences between every other group, based on the order of groups shown in Fig. 1, were significant at beyond the .05 level, e.g., Group PB failed to differ from Groups C or PG but differed from Group PW. Groups PW and C differed at beyond the .01 level.

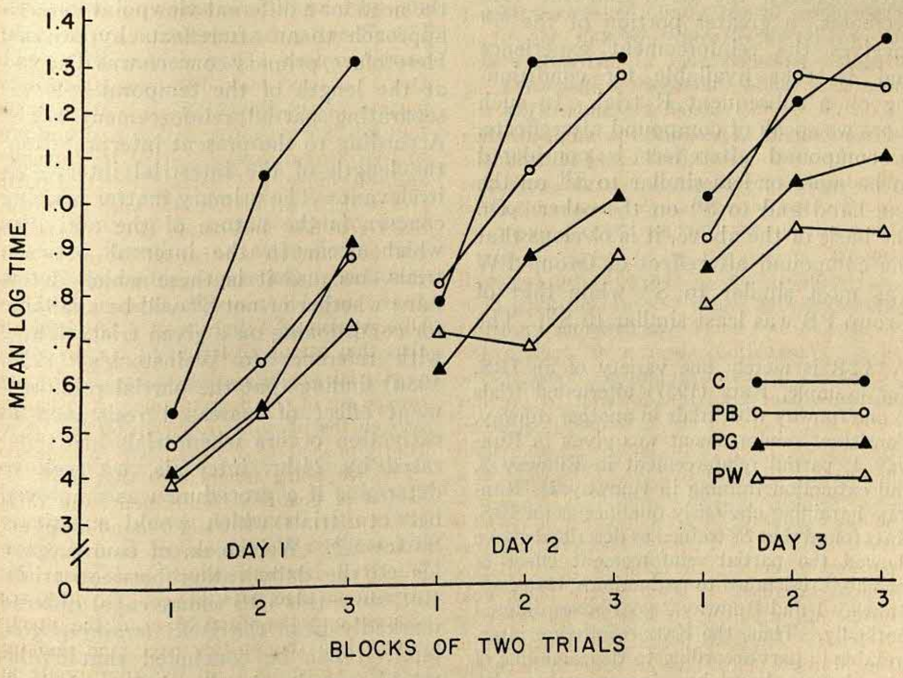


FIG. 1. Running times during extinction.

Differences between adjacent groups, e.g., C vs. PB or PG vs. PW, were not significant, according to Duncan's test, even at the .10 level.

DISCUSSION

The relatively rapid extinction of Group PB confirms results supplied by Capaldi et al. (1963). However, Group PB was slightly more resistant to extinction relative to the C group than was the comparable partial group of the earlier experiment. Such relatively greater resistance may be related to number of acquisition trials, a greater number having been used in this as opposed to the earlier investigation.

Our interpretation of the present results is as follows. It is recognized that under some conditions reinforcement may replace S^N with S^R. The situation in which S^N is occasioned is termed the original nonreinforcement situation (ONS) while the reinforcement situation designed to replace S^N with S^R is termed the interpolated reward situation (IRS).[2] It is assumed that as stimulus similarity between the ONS and the IRS decreases, a greater portion of the S^N survives the reinforcement experience and is thus available for conditioning on a subsequent R trial. In such cases we speak of compound aftereffects. A compound aftereffect is considered to be more or less similar to S^N on the one hand and to S^R on the other. On the basis of the above, it is obvious that the compound aftereffect of Group PW was most similar to S^N while that of Group PB was least similar to S^N. The compound aftereffect of Group PG was presumably of an intermediate similarity.

Considering only the partial reinforcement groups, it can be expected that due to stimulus generalization the S^N occurring in extinction evoked the instrumental response of Group PW with the greatest degree of strength, that of Group PG with the next greatest degree of strength, and that of Group PB with the least degree of strength. The relative degree of resistance expected in the C group will be considered later.

A consideration of the present results in connection with earlier ones (Capaldi & Stanley, 1963) allows a fuller understanding of the implications of this experiment. Capaldi and Stanley presented evidence which suggested that aftereffects remain functional for at least 20 min. On the basis of these data, it was suggested that aftereffects do not undergo diminution, as assumed by Hull and by Sheffield, but rather that a given aftereffect remains functional until replaced by another aftereffect. This new temporal assumption, along with the ITR data presented here and elsewhere (Capaldi et al., 1963), suggests the need for a different viewpoint in one's approach to an aftereffects hypothesis. Heretofore, primary concern was directed at the length of the temporal interval separating partial reinforcement trials. According to the present interpretation, the length of the intertrial interval is irrelevant. The primary matter inviting concern is the nature of the activities which occur in the interval between trials, because it is these which determine whether or not S^N will be available for conditioning on a given trial. Thus, with reference to Weinstock's (1954, 1958) finding that the partial reinforcement effect of increased resistance to extinction occurs when trials are separated by 24-hr. intervals, we seek to determine if a procedure was employed between trials which would substitute S^R for S^N. Weinstock, of course, gave his *S*s the daily ration between trials. But, since this IRS (home cage) differed markedly from the ONS (training goal box), it can be concluded that Weinstock's results hardly infirm—indeed, they support—the present version of the aftereffects hypothesis, although, of course, his results do not support the version of the aftereffects hypothesis offered by Hull and Sheffield.

---

[2] ITR is merely one variety of an IRS. For example, Katz (1957) alternated trials in one runway with trials in another runway. Consistent reinforcement was given in Runway 1, partial reinforcement in Runway 2, and extinction training in Runway 2. Runway 1 training obviously qualifies as an IRS. Katz found that *S*s trained as described above showed the partial reinforcement effect of increased resistance to extinction. However, Runway 1 and Runway 2 goal boxes differed markedly. Thus, the Katz results are interpretable in part according to the same line of reasoning employed here in connection with Groups PW and PG.

For reasons already considered, it is clear that Groups PW and PG are expected to be more resistant to extinction than Group C. However, in comparing Group PB with C the prediction is less straightforward. On the one hand, Group PB could be expected to be no more resistant to extinction than Group C, as was found by Capaldi et al. (1963). Approximately equal resistance would be expected if, in Group PB, ITR completely replaced S^N with S^R. However, as noted, Group PB ran significantly more rapidly than Group C on the initial day of extinction. One interpretation of this finding is that some portion of Group PB's S^N survived the ITR. This, in turn, suggests that the stimulus situation in the IRS and the ONS must have differed in some important respect. One obvious source of stimulus difference would appear to have been the presence of proprioceptive feedback due to locomotion in the ONS and its relative absence in the IRS. Presumably, this difference in stimulus similarity provided for a compound aftereffect in Group PB which bore a mild degree of resemblance to S^N.

That the proprioceptive consequences of the locomotor response contribute a significant portion of the total stimulus complex has long been recognized. However, that such proprioceptive consequences function as aftereffects, as is presently being suggested, is a matter which has received less attention. It can be indicated, first of all, that there seems to be no particular reason for supposing that the stimulus consequences of locomotion fail to function as aftereffects once having admitted that the stimulus consequences of the consummatory response function in such a capacity. Moreover, data presented by Petrinovitch and Bolles (1957) suggest that the stimulus consequences of the instrumental response do indeed function as aftereffects. These investigators demstrated that rats can be trained to turn to the right in a **T** maze for food on a given trial if food was found on the left on the immediately previous trial, and vice versa. Without entering into a full S-R analysis of this behavior, it should be clear that regardless of one's specific interpretation, the stimulus complex evoking the correct response on some current trial must have included the proprioceptive consequences of the turning response executed on the immediately previous trial. Without these proprioceptive consequences, only chance behavior could be expected. Clearly, then, the Petrinovitch and Bolles results strongly suggest that the stimulus consequences of the instrumental response function as aftereffects. Accordingly, the slightly greater resistance to extinction of Group PB relative to Group C in the early trials is understandable in terms of a lack of complete stimulus similarity between the ONS and the IRS.

REFERENCES

CAPALDI, E. J., HART, D., & STANLEY, L. R. Influence of intertrial reinforcement on the aftereffect of nonreinforcement and resistance to extinction. *J. exp. Psychol.*, 1963, **65**, 70–74.

CAPALDI, E. J., & STANLEY, L. R. Temporal properties of reinforcement aftereffects. *J. exp. Psychol.*, 1963, **65**, 169–175.

EDWARDS, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Reinhart, 1960.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KATZ, S. Stimulus aftereffects and the partial reinforcement extinction effect. *J. exp. Psychol.*, 1957, **53**, 167–172.

PETRINOVITCH, L., & BOLLES, R. Delayed alternation: Evidence for symbolic processes in the rat. *J. comp. physiol. Psychol.*, 1957, **50**, 363–365.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. *J. exp. Psychol.*, 1958, **56**, 151–158.

(Received November 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 466–473

DETERMINATION OF ODOR DIFFERENCE LIMENS FOR THREE COMPOUNDS[1]

HERBERT STONE[2]

University of California, Davis

Difference limens and 50% limens were determined at 3 (intensity) I levels with 3 groups of experienced Ss (18 per group) for n-heptyl alcohol, 2-octanone, and ethyl n-valerate. Use of an improved, quantitative olfactometer permitted brief testing periods for 13 test days and results were reported as mg/l air. Data were found to agree with the normal response law and a pooled Weber fraction of 0.23 ± .045 was calculated. An improved formula was developed for calculation of Weber fractions to account for subliminal stimuli. The adjusted Weber fractions were pooled and the K for olfaction was determined to be 0.20 ± .035.

Weber's law states that small equally perceptible increments in response correspond to proportional increments in stimulus, expressed in equation form as: $\Delta S = KS$. This model has been successful in explaining differential sensitivity; that stimulus difference which a panel can detect 50% of the time above chance (when $P = \frac{1}{2}$). Considerable doubt has been cast on application of this law and according to Guilford (1954, pp. 20–42) it is only an approximation to the correct law. Attempts to test responses to olfactory stimuli for agreement with the law have not been successful. Wenzel (1949) noted that there were very few experiments on differential sensitivity and data were inconclusive. Gamble (1898) and Zigler and Holway (1935) used Zwaardemaker's olfactometer which subsequently has been shown to be unreliable for odor testing. Zwaardemaker's data were reported in "olfacties," which are not a quantitative measure. Gamble used a variety of test materials and concluded that responses to olfactory stimuli agreed with Weber's law and the value between $\frac{1}{3}$ and $\frac{1}{4}$. Zigler and Holway used India rubber and reported a Weber fraction which changed from .99 to .17 as the stimulus intensity increased. Wenzel (1949) employed a modified Elsberg apparatus and reported the Weber fraction for phenyl ethyl alcohol as .15. The data were more consistent since concentrations were expressed in molecular terms.

Subsequent investigations have shown the Elsberg apparatus to be unreliable because of the inability of *S*s to perform reliably and the extreme artificiality of the technique (Wenzel, 1955). The experiments reported here employed an olfactometer under controlled temperature, humidity, air flow rate, and odor concentrations which has been shown to be reliable for olfactory testing (Ough & Stone, 1961; Stone, Ough, & Pangborn, 1962). Difference thresholds (ΔI) were determined at three intensities: (I) for three odorous chemicals;

---

[1] This research is based on a dissertation submitted to the Graduate School of the University of California, Davis, in partial fulfillment of the requirements for the PhD degree. The author would like to acknowledge financial support of the Campbell Soup Company. He is indebted to M. A. Amerine and C. S. Ough for their suggestions and assistance throughout this research.

[2] Now at Stanford Research Institute, Menlo Park, California.

2-octanone, *n*-heptyl alcohol, and ethyl *n*-valerate. In addition the 50% limens were determined (that intensity which 50% of the panel can recognize as different from a sample with no odor). Finally, the data were tested for agreement with Weber's law. It was also possible to obtain useful information on the methodology of sample presentation and panel performance.

METHOD

Apparatus.—The olfactometer, described in earlier reports (Stone, 1963; Stone et al., 1962), is based on the premise that air may be saturated to a constant amount by contact with the test material. The odor-saturated air is diluted in a stream of constantly flowing odor-free air and delivered to *S* for his decision.

Calibration.—A previous report indicated the concentration of odorous material could be calculated from knowledge of the experimental temperature, vapor pressure of the test material, and flow rate of air through the sparging apparatus (Stone et al., 1962). Quantitative recovery of different test materials corroborated these calculations (Ough & Stone, 1961; Stone et al., 1962). Additional evidence was provided by quantitative recovery of *n*-heptyl alcohol by means of a gas sampling system connected to a flame-ionization, gas-liquid chromatographic unit. Measurement of peak areas and comparison with a standard curve showed good agreement with the calculated concentrations which were used in the present experiment. A detailed discussion of the recovery procedure may be found elsewhere (Stone, 1963).

Materials.—The three compounds, 2-octanone, *n*-heptyl alcohol, and ethyl *n*-valerate, were chosen for their stability and availability in pure form. The compounds were chemically different and it was of interest to study materials of diverse chemical nature, in order to obtain more general information. The compounds were purified by distillation before each experiment. Prior to the start and after each experiment, samples were checked for purity by a flame-ionization, gas-liquid chromatographic unit. No changes were noted indicating the odorous materials underwent no degradation during testing.

Psychophysical method.—The forced-choice, constant-stimulus method, described by Guilford (1954, pp. 118–153) and modified by Mrak, Amerine, Ough, and Baker (1959), was used in all experiments. Six test samples, evenly spaced above and below the reference concentration, were compared by *S*s to the constant stimuli. Presentation was by a completely randomized, paired design. The task was to indicate the sample within each pair having the more intense odor. The six test samples at each reference level were replicated three times by the 18 *S*s, making a total of 54 determinations for every intensity pair. Table 1 shows the intensity levels and 50% limen levels for all experiments.

Procedure.—All experiments were carried out under carefully controlled conditions. Temperature was set at 24° ± 0.5°C. and the relative humidity, 50 ± 2.5%.

Fifty-four *S*s (40 male and 14 female volunteers), ranging in age from 21 to 50 yr.,

TABLE 1

TEST LEVELS FOR THE THREE COMPOUNDS IN ALL EXPERIMENTS

50% Limen	Intensities × 10⁻⁴ mg/l air		
	n-heptyl alcohol		
3.8	26.6	32.3	36.1
5.7	29.6	36.1	41.9
7.6	30.9	38.0	44.8
9.5	32.3[a]	39.9[a]	47.6[a]
11.4	33.7	41.9	50.5
13.3	35.1	43.8	53.3
	38.0	47.6	58.9
	2-octanone		
1.9	26.6	32.3	36.1
3.8	29.6	36.1	41.9
5.7	30.9	39.0	44.8
7.6	32.3[a]	39.9[a]	47.6[a]
9.4	33.7	41.9	50.5
11.4	35.1	43.8	53.3
	38.0	47.6	58.9
	ethyl *n*-valerate		
1.2	2.7	5.4	13.5
1.6	4.1	8.1	17.6
2.1	4.7	9.5	19.6
2.7	5.4[a]	10.8[a]	21.6[a]
3.8	6.1	12.2	23.7
	6.8	13.5	25.7
	8.1	16.3	29.8

[a] Constant stimulus for each test.

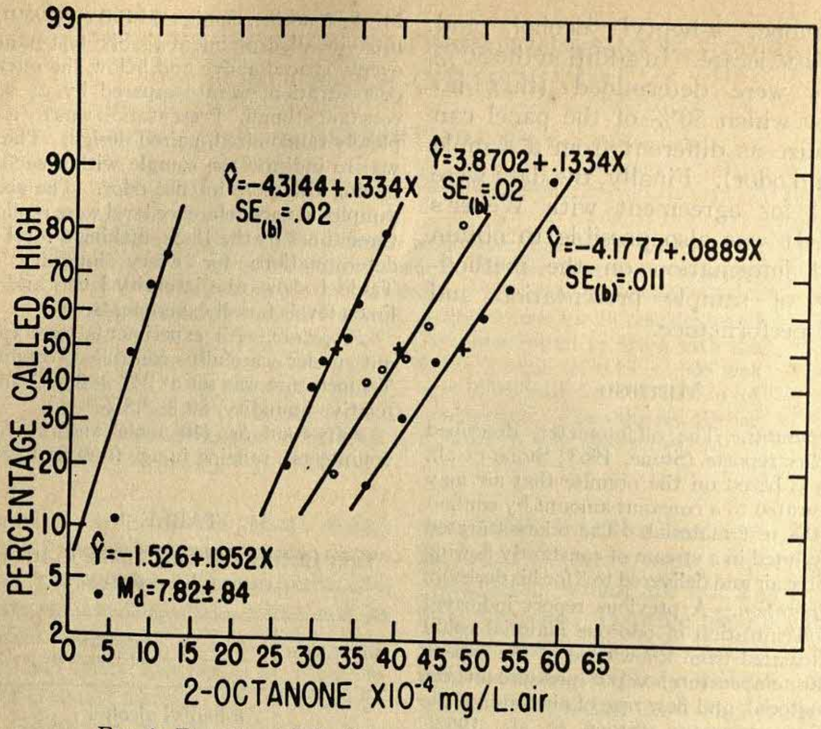


FIG. 1. Equations of the lines, 50% limen, and standard errors of estimate for 2-octanone.

were divided into three equal groups according to their availability for each 13-day test period. Each group was tested with a different compound. All *S*s had previous experience with the apparatus and the test technique.

The *S*s were tested singly the same time of day each test day. The *S* entered the test room and sat down and placed his head in the Plexiglas hood. Seven test pairs were presented and a decision made after each pair. The first pair served as orientation and no decision was necessary. To eliminate conversation a signal system was used: a green light in front of *S* indicated a sample was present. The *S* indicated the more intense sample by pressing a button that lighted a bulb in front of *E*. Each sample was presented for 10 sec., followed by 10 sec. of purified air, with 15-sec. intervals between pairs.

The *S*s were presented with seven test pairs daily and the first pair was always discarded. Preliminary investigations indicated *S*s required the first pair as a "warm up" upon entering the odor-free environment in the hood (Stone et al., 1962). The first day's tests were always repeated on the tenth day to check for learning. No significant differences in response were noted; however, the tenth day was arbitrarily used in the analysis as χ^2 indicated data were more homogeneous.

The 50% limen test was carried out on 3 successive days. The *S*s were presented with seven test pairs daily. One sample of each pair contained no odor; however, *S* knew only that some samples might contain the odor at subliminal intensity.

Treatment of data.—The data were tested for homogeneity by χ^2 analysis daily and at the conclusion of each experiment (Goulden, 1956, pp. 369–372). No deviations from expectation were noted in any experiment and all data were used for the subsequent analysis.

In these analyses, data were not converted to a coded log scale; rather, the linear data were used directly. According to Guilford (1954, pp. 20–42), it is questionable whether any greater precision is attained by use of a log scale. Further, data reported as mg/l air are more meaningful. The maximum-likelihood solution for the constant method was used (Jones, 1957), permitting calculation of *b*, and its reciprocal, $1/b$, which is ΔI at any reference concentration, I. In addition, deviations of the data from the normal response law are noted by a significant χ_R^2 term in the regression analysis.

The test for agreement of the data with Weber's law used a method described by Jones (1957). Proportions of the difference stimuli to the constant stimulus are analyzed by separate maximum-likelihood solutions. The transformed data are necessary so that, regardless of constant stimulus, the regression of the deviates of I would be the same. Insignificant χ_R^2 terms permit the data to be pooled and evaluated by the combined regression analysis. Agreement with the normal response law is noted by insignificant χ_R^2 terms.

RESULTS

The data are reported in the following psychometric terms: ΔI, $\Delta I/I$ (the Weber fraction); $I/I + I_T$ (corrected Weber fraction); b (slope); and $Y = a + bx$ (regression equation). The regression analysis for each I for all compounds is shown in Table 2. There were no significant χ^2 residual terms, indicating agreement of the data with the normal response law. The results for the individual experiments, plotted on normal probability paper as concentration vs. percentage called high, are shown in Fig. 1, 2, and 3. At higher I, the slope of the line increased, indicating a greater I was needed before *S*s could significantly detect a difference in intensity.

Table 3 summarizes all the experimental data including ΔI at each I, Weber fractions, and the 50% thresholds and their standard errors of estimate. In addition, the data from a previous experiment with 2-hepta-

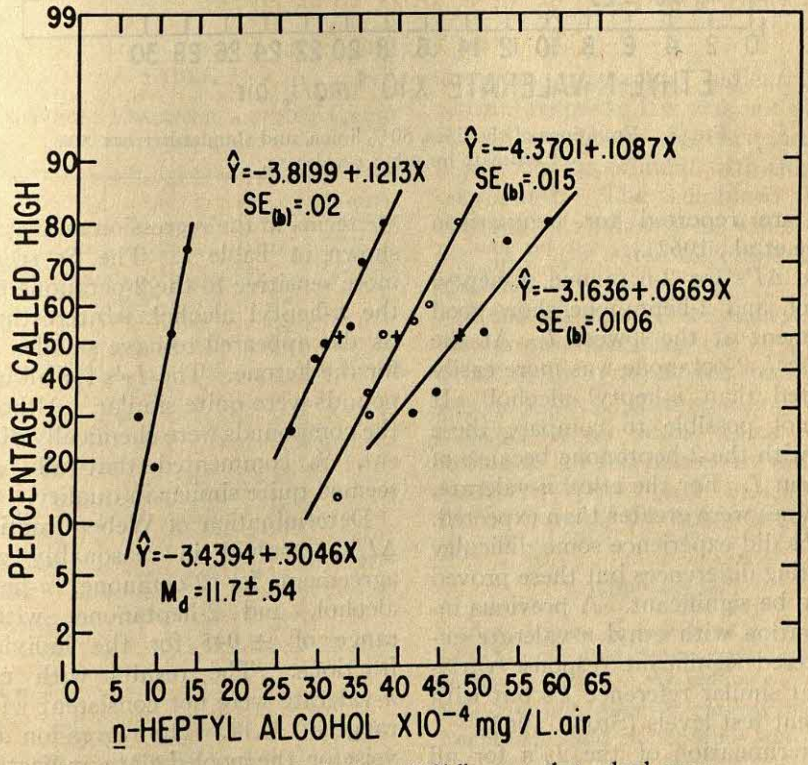


FIG. 2. Equations of the lines, 50% limen, and standard errors of estimate for *n*-heptyl alcohol.

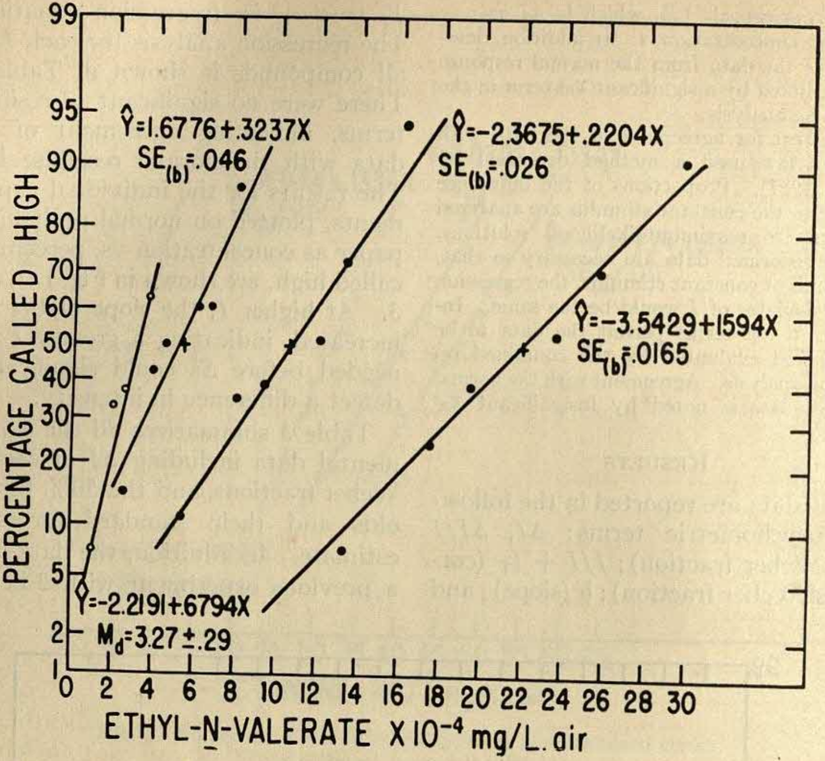


FIG. 3. Equations of the lines, 50% limen, and standard errors of estimate for ethyl *n*-valerate.

none are reported for comparison (Stone et al., 1962).

The ΔI's for 2-octanone, *n*-heptyl alcohol, and 2-heptanone show good agreement at the lowest I. At the highest I, 2-octanone was more easily detected than *n*-heptyl alcohol. It was not possible to compare these data with the 2-heptanone because of different I. For the ethyl *n*-valerate, ΔI values were greater than expected. The *S*s did experience some difficulty detecting differences but these proved not to be significant. A previous investigation with ethyl *n*-valerate encountered significant response variation at similar reference I's but with different test levels (Stone, 1963).

Determination of the I_T's for all compounds resulted in no significant χ_R^2 terms in the regression analyses as shown in Table 2. The *S*s seemed more sensitive to the 2-octanone than the *n*-heptyl alcohol. Interestingly, *S*s also appeared to have smaller ΔI's for the ketone. The I_T's for all compounds were quite similar. Although the compounds were chemically different, *S*s commented that all odors seemed quite similar in quality.

Determination of Weber fractions, $\Delta I/I$, resulted in reasonably good agreement for 2-octanone, *n*-heptyl alcohol, and 2-heptanone, with a range of ±.045 for the individual fractions. The results with ethyl *n*-valerate were not consistent with a range of ±.14. The regression analysis for the pooled data indicated a highly significant χ_R^2 term for the

TABLE 2

Analyses for the Regression of the Concentration for the Three Test Compounds on Percent Correct Response

Intensity $\times 10^{-4}$ mg/l air	*SS*s Total ($df = 5$)	Regression ($df = 1$)	Residual ($df = 4$)	χ_R^2	p
2-octanone					
32.3	0.8050	.7847	.0203	1.0960	$.80 > p > .90$
39.9	0.8008	.6984	.1024	5.5290	$.20 > p > .30$
47.6	1.2320	1.1971	.0349	1.8840	$.70 > p > .80$
50% limen	0.2622	.2329	.0293	1.5820	$.30 > p > .90$
n-heptyl alcohol					
32.3	0.6930	.6626	.0304	1.642	$.80 > p > .90$
39.9	1.0540	.9518	.1022	5.519	$.20 > p > .30$
47.6	0.9010	.7329	.1681	9.077	$.05 > p > .10$
50% limen	0.3046	.2589	.0457	2.468	$.50 > p > .70$
ethyl *n*-valerate					
5.4	1.0436	.9208	.1228	6.631	$.10 > p > .20$
10.8	1.4592	1.3440	.1152	6.221	$.10 > p > .20$
21.6	1.7619	1.7352	.0267	1.442	$.80 > p > .90$
50% limen	0.2257	.1804	.0453	2.446	$.50 > p > .70$

TABLE 3

ΔI's, Weber Fractions, and 50% Limens for the Test Compounds

I $\times 10^{-4}$ mg/l air	ΔI	K[a]	K_1[b]	$I_T \pm SE$[c] $\times 10^{-4}$ mg/l air
2-octanone				$7.82 \pm .84$
32.3	7.49	0.23	0.19	
39.9	10.5	.26	.22	
47.6	11.25	.24	.20	
Average		.24	.20	
n-heptyl alcohol				$11.17 \pm .54$
32.3	8.24	0.25	0.19	
39.9	9.19	.23	.18	
47.6	14.95	.31	.25	
Average		.27	.21	
ethyl *n*-valerate				$3.27 \pm .29$
5.4	3.1	0.57	0.36	
10.8	4.54	.42	.32	
21.6	6.3	.29	.25	
Average		.43	.31	
2-heptanone[d]				8.97 ± 1.09
31.5	7.69	0.24	0.19	
85.3	19.75	.23	.21	
112.5	25.24	.22	.21	
Average		.23	.20	

[a] $K = \frac{\Delta I}{I}$.

[b] $K_1 = \frac{\Delta I}{I + I_r}$, $I_r = 50\%$ limen.

[c] $I_T \pm SE = 50\%$ limen ± standard error of estimate.

[d] Data from Stone et al. (1962).

common regression, indicating the normal response law was not obeyed. A similar problem was encountered in a previous experiment with this same compound. The combined Weber fraction for all data is .28 which would be .23 if the ester experiment were excluded.

Weber fractions determined from the formula including the 50% limen would seem to adjust the fractions to smaller values. The resulting Weber fraction for all odors tested is .23, similar to .28 determined by the usual method. These data indicate the Weber fraction for odors is approximately .2.

Discussion

Comparison of these data with the 2-heptanone data shows good agreement at the lowest I. At the other I, the ΔI's are not directly comparable; the 2-octanone and *n*-heptanol show reasonably close agreement, although the panel was more sensitive to the ketone, a fact borne out by the lower 50% limen. It would be of interest to study other ketones and alcohols to determine the exact pattern of sensitivity for these compounds.

Determination of Weber fractions resulted in good agreement between all the compounds. Wenzel (1949) reported the Weber fraction for phenyl ethyl alcohol as .15 as compared with .23 reported here. The *S*s were not tested by the same test method and the technique was different; however, agreement is close. There has been considerable effort expended in attempting to improve on Weber's law which is acknowledged to be only the first approximation to such a law (Guilford, 1954, pp. 20–42). At the concentration extremes, the law has not been found to hold for a variety of reasons. While the present experiments did not cover the entire range of possible intensities it was approached. Most *S*s noted that the most intense samples were "almost too strong." Stevens (1957) and others have suggested a power law for subjective intensity and for odors, the value *n* was found to be .55–.6. Jones (1958a, 1958b) has demonstrated that this power law holds quite well for many odors of diverse chemical nature. Others have suggested modifying Weber's law to account for the discrepancies reported (Stevens, 1951, pp. 26–36; Stewart, 1962).

Use of the K_1 formula for determination of corrected Weber fractions was developed from a suggestion of Stevens (1957) and Schutz (1955), originally proposed by Miller (1947).

The difference limen (ΔI) is a function of the stimulus (I) and depends on activation of discrete neural units, a process shown to be of an all-or-none character (Allison & Warwick, 1949). According to Miller and Garner (1944) and Miller (1947), in the activation of these neural units, the stimulus is not only the exact amount required just to evoke a response; but is usually more. This surplus of stimuli will contribute to *S*'s response and therefore influence the ΔI, but will vary at any time as the overall sensitivity varies. There is no way of preventing this residual from influencing response and it would be difficult to measure. It is assumed that while *S* cannot detect below I_T; nevertheless, this stimulus can influence response. This may be considered part of the interfering stimuli; i.e., changes in neural sensitivity, shifts in attention, variability of equipment, and other factors which become more significant as the 50% limen is approached. To adjust the Weber fraction to account for these discrepancies, Miller (1947) proposed the following empirical equation:

$$K = \frac{\Delta I}{I + I_r} \qquad [1]$$

where I_r is some small addition to I. It is an additive term since it influences response positively; i.e., there is something present but *S* cannot identify it as a true stimulus.. Obviously, this I_r must be some value below I_T, the 50% limen. In the present experiments it has arbitrarily been assumed that $I_r = I_T$. Therefore, the proposed Weber fraction now reads:

$$K_1 = \frac{\Delta I}{I + I_T} \qquad [2]$$

Application of the data to Equation 2 permitted calculation of adjusted Weber fractions listed as K_1 values in Table 3. The results for 2-octanone, *n*-heptyl alcohol, and 2-heptanone were very similar with an average Weber fraction of .20 ± .035. The data for ethyl *n*-valerate did not yield a similar value; however, there was a decrease in the Weber fraction from .43 to .31. This lack of agreement with other test materials was attributed earlier to the significant χ_R^2 terms reported in the regression analyses and would apply here as well. The combined Weber fraction for all test materials including ethyl *n*-valerate is .25, still reasonably close to the .15 for phenyl ethyl alcohol reported by Wenzel (1949). Better agreement is noted if the ethyl *n*-valerate is omitted from these results. While the compounds of the present experiment were similar in odor quality and structure, the panels were all different, albeit a university population. Variations noted may be attributed to the few replications per judge per level of intensity. The data of the ethyl-valerate study indicated the panel required additional replications as their differential sensitivity was not as good, especially at the lower intensities. Interestingly, *S*s recorded their lowest 50% limen with this ester despite their difficulty in detecting differences.

No further conclusions are possible until additional quantitative experiments on differential sensitivity are conducted with odors of diverse chemical and subjective nature. These data do suggest a possible approach to Weber's law as it applies to olfaction. It is conceivable that future experimentation will indicate I_r to be $\frac{1}{2}$ I_T, a sort of average value for the surplus stimuli. However, for the present experiment I_T was arbitrarily chosen as the correction term. Experimentation at or near the 50% limen and up to the trigeminal level should provide the solution to some of these problems.

REFERENCES

ALLISON, A. C., & WARWICK, R. T. Quantitative observations on the olfactory system of the brain. *Brain*, 1949, **72,** 186–197.

GAMBLE, E. A., McC. The application of Weber's law to smell. *Amer. J. Psychol.*, 1898, **10,** 82–142.

GOULDEN, C. H. *Methods of statistical analysis*. New York: Wiley, 1956.

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

JONES, F. N. Scales of subjective intensity for odors of diverse chemical nature. *Amer. J. Psychol.*, 1958, **71,** 305–310. (a)

JONES, F. N. Subjective scales of intensity for three odors. *Amer. J. Psychol.*, 1958, **71,** 423–425. (b)

JONES, L. V. Methodology of preference measurements. *USA Qm. Food Container Inst. res. Rep.*, 1957, No. 4, 12–51.

MILLER, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. Acoust. Soc. Amer.*, 1947, **19,** 609–619.

MILLER, G. A., & GARNER, W. R. Effect of random presentation on the psychophysical function: Implications for a quantal theory of discrimination. *Amer. J. Psychol.*, 1944, **57,** 451–467.

MRAK, V., AMERINE, M. A., OUGH, C. S., & BAKER, G. A. Odor difference test with applications to consumer preferences. *Food Res.*, 1959, **24,** 574–578.

OUGH, C. S., & STONE, H. An olfactometer for rapid and critical odor measurement. *J. food Sci.*, 1961, **26,** 452–456.

SCHUTZ, H. G. Differential sensitivity in gustation. Unpublished doctoral dissertation, Illinois Institute of Technology, 1955.

STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 26–36.

STEVENS, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64,** 153–181.

STEWART, J. L. A law for loudness discrimination. *Science*, 1962, **137,** 618–619.

STONE, H. Some factors affecting olfactory sensitivity and odor intensity. Unpublished doctoral dissertation, University of California, Davis, 1963.

STONE, H., OUGH, C. S., & PANGBORN, R. M. Determination of odor difference thresholds. *J. food Sci.*, 1962, **27,** 197–202.

WENZEL, B. M. Differential sensitivity in olfaction. *J. exp. Psychol.*, 1949, **39,** 129–143.

WENZEL, B. M. Olfactometric method utilizing natural breathing in an odor-free "environment." *Science*, 1955, **121,** 802–803.

ZIGLER, M. J., & HOLWAY, A. H. Differential sensitivity as determined by amount of olfactory substance. *J. gen. Psychol.*, 1935, **12,** 372–382.

(Received November 28, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 474–484

RELATIONSHIPS BETWEEN CHOICE TIME AND FREQUENCY DURING DISCRIMINATION TRAINING AND GENERALIZATION TESTS[1]

KEITH A. WOLLEN[2]

University of Iowa

80 human Ss were given 60 discrimination training trials in which R_1 was always correct for S_1, but confusion was induced on S_2 by having R_2 correct on 40, 60, 80, or 100% of the trials and R_1 correct on the rest. Subsequently, a total of 105 test trials was given in random order on S_1, S_2, and 5 intermediate stimuli. The results were: (a) choice time lengthened as the relative frequency of response approached 50%; (b) points of maximum time and the 50% points shifted toward S_2 as the percentage of trials that R_2 was correct for S_2 decreased from 100 to 40%; (c) the time of each response increased as its frequency decreased, with the rate of increase being greater for slow than for fast responders.

One purpose of the present experiment was to determine how the choice time between two incompatible responses is affected by variations in the relative strengths of the two alternatives. The task was to learn which of two responses, R_1 or R_2, should be given to each of two stimuli, S_1 and S_2. In one condition, R_1 was always correct for S_1 and R_2 was always correct for S_2. However, in other conditions, confusion was introduced on S_2 trials by designating R_2 as correct on some of these trials and R_1 on the remainder. Subsequent to this training, generalization tests were administered to determine the frequency with which each of the responses would be elicited by the training and intermediate stimuli.

The training phase of the experiment may be viewed as establishing three response tendencies: (*a*) R_2 to S_2, (*b*) R_1 to S_2, and (*c*) R_1 to S_1. Each of these tendencies would be assumed to generalize to other similar stimuli. Since the second and third tendencies involve the same response, they may be combined into a net tendency to make R_1. Consequently, there would be a tendency to elicit R_1 which would be strongest at S_1 and generalize toward S_2, as well as a tendency to make R_2 which would be strongest at S_2 and generalize toward S_1. Therefore, as the stimulus became more similar to S_2, the tendency to make R_1 would decrease whereas the tendency to make R_2 would increase. Presumably, the two tendencies would be equal at some stimulus value between S_1 and S_2. The question under consideration is whether choice time would increase as the two tendencies became more nearly equal, and be maximal when they were in fact equal. Since the relative frequency of response was used to assess the strengths of the tendencies, the above question becomes one of determining whether choice time becomes maximal as the percentage frequency of response approaches 50%.

Several *E*s (Berlyne, 1957; Cartwright, 1941; Festinger, 1943; Johnson, 1939; Kellogg, 1931; Lemmon, 1927) have found that choice time increased as the relative frequency of the responses approached 50%. However, each of these studies utilized some version of the psychophysical method of constant stimuli in which *S*s were instructed to make one response to one stimulus and a second response to a second stimulus. A purpose of the present experiment was to see whether a similar finding would obtain when the response tendencies were established by training rather than by instruction.

---

[1] This paper is based upon an MA thesis submitted to the University of Iowa. The author is indebted to Harold P. Bechtoldt for his advice and assistance through the course of this investigation.

[2] Now at Washington State University, Pullman, Washington.

A second purpose of this study was to determine whether the stimulus value at which the two response tendencies were equal (i.e., the point of 50% relative frequency) would be shifted toward S_2 by increasing the tendency to make R_1 and decreasing the tendency to make R_2. As the percentage of times that R_2 is correct on S_2 trials decreases, the generalized tendency to make R_2 should also decrease but the R_1 tendency should increase. Thus curves depicting the strength of each of the two tendencies should cross nearer to S_2 and the points of 50% relative frequency should, of course, likewise be displaced toward S_2.

In addition to an investigation of the relationship between frequency and choice time irrespective of the particular choice made, a further goal of this study was to examine the relationship between the frequency and time of R_1 choices separately from those of R_2. The major reason for this interst is the assumption of several theorists (e.g., Hull, 1952; Spence, 1956) that response time is inversely related to response frequency. Although these theories were not formulated to account for human learning situations of the kind described herein, it seems most reasonable to expect weak tendencies to be associated with lower frequencies and longer times than would be the case for strong tendencies. In other words, as the stimuli become more similar to S_2, the frequency of R_1 should decrease and the response time should increase. Similarly, as the stimuli become more similar to S_1, the frequency of R_2 should decrease and the response time should increase.

The data relevant to the hypothesis of an inverse relation between time and frequency are somewhat equivocal. Studies in which *S*s were instructed to respond to the training stimulus but to inhibit responses to other stimuli have not generally found such a relationship (Brown, Bilodeau, & Baron, 1951; Gibson, 1939; Mednick, 1955). However, as Gibson suggested, such a failure may very likely be due to the fact that *S*s inhibit potential long latency responses to stimuli other than the training stimulus. Mednick (1958) attempted to reduce such a tendency by giving *S*s a set to respond very quickly and reported an inverse relationship, but only for the fast responders.

Another way to eliminate the tendency for *S*s to inhibit potential long latency responses is to require them to make a response on each trial. Those using this technique have generally found an inverse relationship (Buss, 1950; Kellogg, 1931; LaBerge, 1961), although the latter two *E*s found some tendency for the relationship to break down for responses to stimuli most distant from the training stimulus. The present investigation will likewise use the required response paradigm in an attempt to assess the reliability of this breakdown.

METHOD

Apparatus.—The apparatus consisted of three main components, a stimulus unit, a response unit, and *E*'s control unit. The stimulus unit consisted of a vertical line of light $\frac{1}{32} \times \frac{3}{8}$ in. which could be moved to different horizontal positions between two white reference spots, each $2\frac{1}{4}$ in. square. The squares were placed so that the lower inside corners were 4 in. above and 5 in. to either side of the stimulus when the latter was in the center position. The entire area around the stimulus and squares was painted a flat black. A well shielded 60-w. lamp placed above the stimulus unit provided just enough light to enable *S* to see the reference squares clearly after a period of dark adaptation. However, the level of illumination was too low to be measured by a MacBeth illuminometer.

The response unit consisted of two spring-loaded levers mounted on a table in such a manner that *S* could push them only toward or away from himself. A black, cloth tunnel extended $7\frac{1}{2}$ ft. from the stimulus unit and ended in a headrest for *S* at the response end of the apparatus.

The *E*'s control unit, which was located behind the stimulus panel, consisted of devices which enabled *E* to position and present the stimulus and appropriate informing signals. A small magnifying glass mounted on the apparatus enabled *E* to locate the stimulus at different positions with less than .1 mm. error. Hunter timers controlled the duration of the stimulus exposure, the intertrial interval, and the informing signals. Choice times were recorded by means of a Standard Electric clock. A 12-in. electric fan served to mask the noise of the apparatus.

Procedure.—The *S* was seated and his head clamped loosely in the headrest. Then all lights were turned out for a 10-min. period of dark adaptation. At the end of this time, the shielded 60-w. lamp was turned on and *S* was asked to specify the shape of the reference squares. If he could not do this, more dark adaptation was given. Once the criterion was satisfied, the 60-w. lamp was left on for the remainder of the experiment. For all phases of the experiment, *S* was instructed to fixate between and a little below the squares which were then in the peripheral field of vision. The reason for this was that some pilot *S*s reported the squares sometimes appeared to fade out when foveal vision was employed.

After the instructions had been read, *S* was given 60 trials to learn which response, R_1 or R_2, to make to each of two stimuli, S_1 and S_2. Each trial consisted of a 2-sec. warning tone, presentation of the stimulus light for 4 sec. or until *S* responded, and a 4-sec. intertrial interval. Of the 60 trials, 30 were with the stimulus in Position S_1 (15 mm. to one side of the center position) and 30 were with the stimulus in Position S_2 (15 mm. to the opposite side of center). The *S*'s task was to push the response lever on the side of his preferred hand either forward or backward, depending upon which response he thought would be correct. Each response was followed immediately by the termination of the light, and a chime if correct or a buzzer if not.

The instructions emphasized the appearance of different stimuli, the necessity for making only one response per trial, the relatively long time available for responding, and the importance of accuracy rather than speed. The *S*s were further instructed to base their judgments upon the horizontal position of the light, and warned that any attempt to find a "solution" in terms of the pattern of forward and backward responses would be unsuccessful inasmuch as there was no systematic pattern.

Although R_1 was always correct for S_1, it was correct for S_2 only 0, 20, 40, or 60% of the time. Each of the resulting four groups is designated by two numbers, the first representing the percentage of trials on which R_1 was correct for S_1, and the second the percentage of times R_1 was correct for S_2. Thus, for Group 100-0, R_1 was correct on 100% of the S_1 trials and 0% of the S_2 trials; R_2 was correct on all S_2 presentations. For Group 100-20, R_1 was correct for 100% of the S_1 trials as well as 20% of the S_2 trials, but R_2 was correct on the remaining 80% of the S_2 trials. Groups 100-40 and 100-60 complete the four conditions. The side, right or left, on which S_1 appeared, and the response, forward or backward, designated correct for that side, were counterbalanced so that there were four subgroups within each of the four main groups.

The order of presentation of the two stimuli was random with the restrictions that Trials 1 and 60 would be to S_1 and that each block of 10 training trials would consist of 5 to S_1 and 5 to S_2. Although the order of presentation of S_1 and S_2 was identical for all *S*s, the reinforcement sequences necessarily varied since the proportion of trials on which R_2 was correct for S_2 differed from group to group.

Immediately following the 60 training trials, all *S*s were given a series of 105 test trials which included 15 trials each on S_1, S_2, and five intermediate positions. The test trials were identical in all respects to the learning trials except for the omission of the informing signals. All *S*s had been informed in the instructions that they would have a series of learning trials followed by a series of tests and that the informing signals would not be used during the testing phase.

The sequence of stimulus positions, which was identical for all *S*s, was essentially random except for the restrictions that each block of seven tests would contain one trial to each of the seven positions, that the stimulus would be in the center position on the first trial of each of the first three blocks, and that the stimulus would be symmetrical about the center on the second and third trials of each of the first three blocks. These latter two restrictions were imposed merely to facilitate comparison of these results with data as yet unpublished.

After the experimental session, each *S* was asked to indicate verbally the number of positions he though was present in testing and the correct response for each.

Subjects.—The *S*s were 103 students from an introductory psychology course. Those who did not meet a visual acuity criterion of no more than one error on the 20/25 line of a Snellen Eye Chart were disqualified from participation. Since the primary interest was in the comparison of data from *S*s who based their judgments on the horizontal position of the stimulus, *S*s who used other criteria were discarded. Thus 3 *S*s were eliminated because they experienced a marked autokinetic phenomenon and 4 because they based their decisions upon the vertical position of the stimulus which did not change. In addition, it was desired to use only those *S*s who had learned the task sufficiently well. Therefore, 9 *S*s were discarded for failure to meet a learning criterion of no more than two failures to give R_1 in the last 10 S_1 trials, and 7 were eliminated who indicated in their verbal solutions that the object was to make R_2 to S_1. Each of these *S*s was rejected immediately and the next *S* assigned to the same condition. Approximately the same number of *S*s was disqualified in each of the four groups. A total of 20 *S*s per group entered into the final analysis.

RESULTS AND DISCUSSION

Acquisition

Frequency.—The percentage frequency data from acquisition are presented in Fig. 1. As is apparent, the relative frequency of each response to each stimulus roughly approximated the percentage of trials that the response was correct for that stimulus; i.e., the relative frequency of R_1 to S_1 approached 100% for all groups, whereas the frequency of R_2 to S_2 roughly corresponded to the proportion of trials that R_2 was cor-

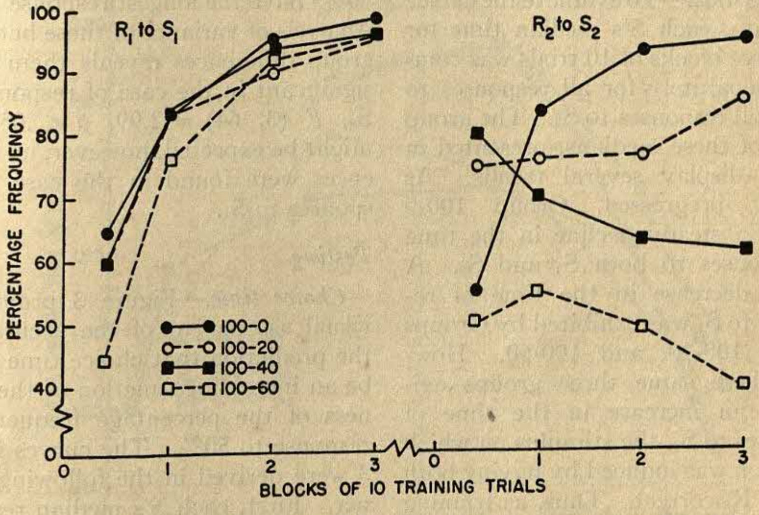


FIG. 1. Percentage frequency of response of each group to each training stimulus as a function of trials. (The first point on each curve represents the frequency on the first trial, whereas subsequent points are means of blocks of 10 trials.)

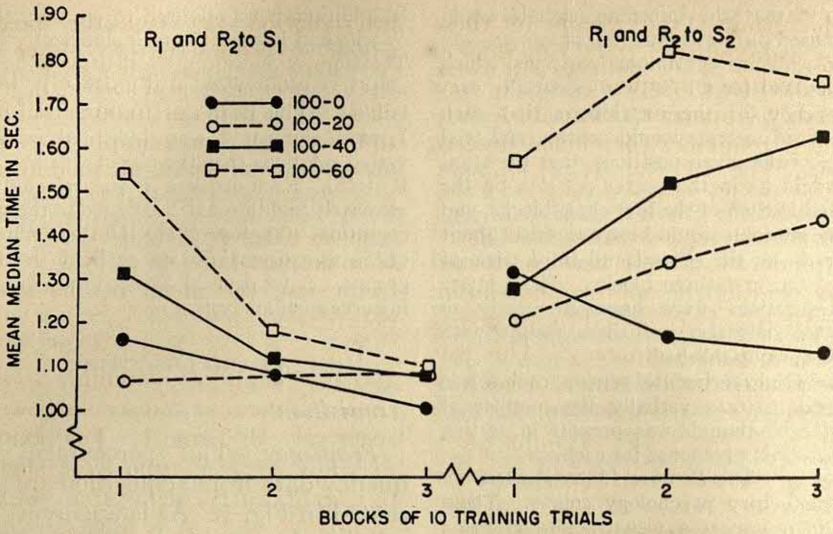


FIG. 2. Response time for each group on each training stimulus as a function of trials.

rect for S_2. Analysis of variance on the last block of 10 trials revealed no statistically significant differences in the case of the R_1 responses, but the differences among the R_2 responses were significant, $F(3, 64) = 49.14$, $p < .001$.

Choice time.—To evaluate the choice time data, each *S*'s median time for successive blocks of 10 trials was computed separately for all responses to S_1 and all responses to S_2. The group means of these medians, presented in Fig. 2, display several trends. As training progressed, Group 100-0 showed a steady decline in the time of responses to both S_1 and S_2. A similar decrease in the time of responses to S_1 was exhibited by Groups 100-20, 100-40, and 100-60. However, these same three groups evidenced an increase in the time of responses to S_2, the stimulus on which confusion was induced by having both R_1 and R_2 correct. Thus, as training proceeded, choice time became progressively longer if the stimulus was associated with confusion, but shorter if little or no confusion was present.

A similar relationship between conflict and choice time is found in a between-groups comparison. On the third block of trials, the groups with the greatest amount of confusion (defined in terms of the nearness of the relative frequency of response to 50%) have the longest response times. Analysis of variance of these between-group differences reveals them to be significant in the case of responses to S_2, $F(3, 64) = 2.99$, $p < .05$. As might be expected, however, no differences were found in the case of responses to S_1.

Testing

Choice time.—Figure 3 provides a visual assessment of the accuracy of the prediction that choice time would be an increasing function of the nearness of the percentage frequency of response to 50%. The curves in Fig. 3 were derived in the following manner. First, each *S*'s median response time (considering R_1 and R_2 together) and percentage frequency of R_1 were determined for each of the seven stimulus positions. Then, for each of the groups, the means of these times were plotted as a function of the percentage frequency of R_1. As is readily apparent, there is a pronounced tendency for choice time to increase as the relative frequency of R_1 approaches 50%.

To provide a statistical analysis of this tendency, a test was conducted to see whether the points of maximum response time correspond to the points of 50% relative frequency. This was accomplished by plotting, for each *S*, the percentage frequency of R_1 as a function of the seven stimulus positions. For the majority of *S*s, this resulted in a curve which was near 100% at S_1, decreased gradually as it approached S_2, and at some point fell below 50%. Then the stimulus position (in terms of distance in millimeters from S_1) at which each *S*'s R_1 frequency curve crossed 50% was found. Since few *S*s had exactly 50% R_1 at any of the positions, it was necessary in most cases to use linear interpolation between the position which was immediately above and that which was immediately below 50%. The point of maximum choice time for each *S* was simply the stimulus position at which his median time was longest. For the few cases in which there was more than one point of maximum time or of 50% relative frequency, the mean of the several points was used.

The group means of the 50% points and the points of maximum time are shown in the first and second rows, respectively, of Table 1. The points of maximum choice time were slightly, but significantly, higher than the 50% points, $F\ (1,\ 76) = 6.18$, $p < .025$. These data appear contrary to those of others who found that the points of maximum time correspond to the 50% points (Berlyne, 1957; Cartwright, 1941; Festinger, 1943; Johnson, 1939; Kellogg, 1931; Lemmon, 1927). However, although the discrepancy was statistically significant, it was rela-

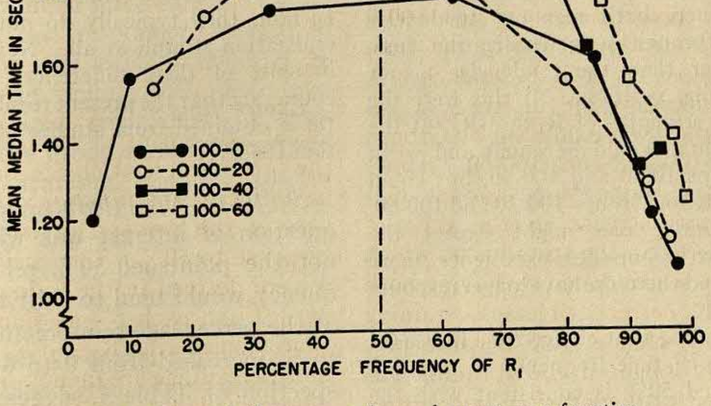


FIG. 3. Time of all responses for each group as a function of the relative frequency of R_1.

TABLE 1

MEANS OF THE POINTS OF 50% RELATIVE FREQUENCY AND THE POINTS OF MAXIMUM RESPONSE TIME FOR EACH GROUP IN TESTING

Point	Group				M
	100-0	100-20	100-40	100-60	
50% frequency	16.92	18.39	21.37	24.46	20.28
Maximum time	17.75	19.25	23.52	25.88	21.60

tively small in comparison to the differences among groups. Moreover, the difference was reduced from 1.32 to 0.56 mm. when each *S*'s maximum mean time was employed in lieu of his maximum median time. Thus it seems reasonably safe to say that choice time reached a maximum at a point which corresponded very closely with the 50% points.

One feature of the data in Fig. 3 merits special attention, namely, the fact that choice times in the 50% region are longer for Groups 100-40 and 100-60 than for the other two groups. One possible ad hoc "explanation" of this is that the point of 50% response occurs near S_2 for Groups 100-40 and 100-60. Hence, both R_1 and R_2 would be relatively strong since S_2 was a training stimulus for both responses. In contrast, the 50% points for Groups 100-0 and 100-20 fall nearer to the center position. As a consequence, both response tendencies might, through a generalization loss, be weaker than the tendencies nearer the training positions. If this were the case, the strengths of R_1 and R_2 (at the 50% points) would be equal and weak for Groups 100-0 and 100-20, but equal and strong for Groups 100-40 and 100-60. Consequently, one might expect the latter two groups to experience more conflict and therefore have longer response times.

The finding that choice time increased as the percentage frequency of response approached 50% is consistent with the findings of those working within the framework of information theory (e.g., Brainard, Irby, Fitts, & Alluisi, 1962; Crossman, 1953, 1955; Crossman & Szafran, 1956; Fitts, Peterson, & Wolpe, 1963; Hyman, 1953). These *E*s have found that reaction time increased with increases in uncertainty. In the present study, uncertainty would be assumed to increase as the relative frequency approached 50%. Therefore, reaction time should also increase as was in fact the case. However, it should be noted that the present study differs in several important respects from those conducted by the information theorists. In studies of information theory, "uncertainty" refers primarily to the probability of a stimulus occurring on a particular trial. In contrast, uncertainty in the context of the present experiment refers to the probability of a response being correct for a particular stimulus, and hence the probability of the response rather than of the stimulus. Associated with this difference is the fact that information theorists designate one and only one response as correct for a particular stimulus, whereas two responses were correct for Stimulus S_2 in the present study. Moreover, information theorists employ well practiced *S*s who are instructed to respond as quickly as possible, whereas the present study utilized relatively naive *S*s who were told to take their time. Finally, studies on information theory have not placed primary emphasis upon times obtained on generalization tests as was the case herein. In fact, they typically do not use generalization stimuli at all. Nevertheless, in spite of these differences, it is encouraging that the present results parallel those obtained from studies of information theory.

Shifts in 50% points.—A second question of interest was whether or not the points of 50% relative frequency would tend to shift toward S_2 as the percentage reinforcement of R_1 to S_2 increased from 0 to 60%. Inspection of Table 1 suggests such a trend and an analysis of variance substantiates its significance, $F(3, 76) = 16.80$, $p < .001$. As might be expected, the points of maximum choice time displayed shifts analogous to those in response frequency.

Relationship of frequency and time.—To investigate the possibility that the time of a response is inversely related to its frequency, each *S*'s median time of R_1 responses only was found for each stimulus position. Medians based only on one response were excluded from consideration. The means (for each group) of these medians were then examined to see whether there was a progressive lengthening of response time as the frequency of R_1 decreased, i.e., as the test stimulus moved further away from S_1 and toward S_2. A similar procedure was followed for the R_2 data. The results of this analysis showed that response time tended to be inversely related to response frequency, but the relationship tended to break down for responses to stimuli most distant from the training stimulus.

Since the inverse relationship seemed more evident for individual *S*s than for grouped data, further analysis was undertaken. It was noted that some *S*s responded quite rapidly at all stimulus positions whereas others responded much more slowly. Moreover, the fast responders tended to give R_1 in the neighborhood of S_2 and R_2 in the vicinity of S_1, more than did the slow responders. Consequently, the group means for stimuli near the training stimulus tended to be composed of both fast and slow responders, whereas the means for positions farthest from the training stimulus were composed primarily of fast responders. Hence, when the times for all *S*s within a given group were combined, the net result was a reduction in the times of responses to stimuli most distant from the training positions. This suggests that the partial failure to obtain the expected relationship may have been due, at least in part, to an artifact of the combination of data from fast and slow responders.

As a check on this hypothesis, the *S*s within each group were divided into the 10 slowest and the 10 fastest on the basis of each *S*'s mean response time on all test trials at all stimulus positions. Then the mean median response time in seconds for each position was found separately for the fast and slow responders. For each position, the mean was based only on those *S*s who gave at least two responses to that stimulus. The resulting values are presented in the left-hand portion of Table 2. The number below each response time indicates the percentage of trials that R_1 was given when that particular stimulus was presented. These percentages were obtained by dividing the total number of R_1 responses to each position by the total number of responses (R_1 plus R_2) made by these *S*s. This was done separately for the slow and fast responders. The analogous times and percentages of R_2 are shown in the right-hand portion of Table 2. The four blank entries in the R_2 data indicate that no *S* gave R_2 two or more times at those positions. For two of these blank entries there is a percentage response given. This merely reflects the fact that single responses were considered in the frequency data but not in computing the mean times.

As these data attest, there was a pronounced tendency for response time to increase as frequency decreased. Moreover, the rate of increase was considerably greater for the slow responders. This fact, together with the finding that re-

TABLE 2

TIME AND PERCENTAGE FREQUENCY IN TESTING OF R_1 AND R_2 AT EACH STIMULUS POSITION FOR FAST AND SLOW *S*s

*S*s	Group	R₁							R₂						
		Stimulus Position							Stimulus Position						
		S₁						S₂	S₁						S₂
		0	5	10	15	20	25	30	0	5	10	15	20	25	30
Slow	100-0	1.26 100	1.44 99	1.94 92	2.17 76	2.38 41	2.47 10	(2.70) 5	— 0	— 1	2.40 8	2.49 24	1.96 59	2.06 90	1.48 95
	100-20	1.34 98	1.58 95	1.89 81	2.31 65	2.66 40	2.99 27	2.84 17	(3.20) 2	(1.72) 5	2.72 19	2.64 35	2.58 60	2.20 73	2.04 83
	100-40	1.67 99	1.46 94	2.02 91	2.56 80	2.76 67	2.99 38	2.81 27	(3.15) 1	(3.00) 6	3.32 9	2.41 20	2.56 33	2.43 62	2.31 73
	100-60	1.58 98	1.78 98	2.05 88	2.15 83	2.78 72	2.98 60	2.95 40	(2.60) 2	— 2	2.95 12	2.72 17	2.81 28	2.75 40	2.64 60
Fast	100-0	0.87 95	0.90 87	1.20 77	1.32 47	1.28 24	1.01 9	(1.00) 3	(1.25) 5	1.21 13	1.49 23	1.39 53	1.30 76	1.00 91	0.83 97
	100-20	0.88 95	0.96 91	1.14 77	1.37 61	1.14 32	1.28 17	1.28 10	(0.70) 5	1.05 9	1.61 23	1.47 39	1.30 68	1.23 83	0.96 90
	100-40	1.20 91	1.09 89	1.18 75	1.18 74	1.43 59	1.56 55	1.55 29	(1.90) 9	1.42 11	1.44 25	1.76 26	1.60 41	1.54 45	1.49 71
	100-60	0.91 100	1.02 96	1.03 92	1.22 87	1.46 65	1.30 57	1.68 48	— 0	(1.50) 4	1.71 8	1.76 13	1.78 35	1.60 43	1.74 52

grouping the data into fast and slow categories greatly reduced the number of exceptions to an inverse relationship, lend support to the previous contention that the partial failure to obtain such a relationship was probably due to an artifact of combining data from *S*s who respond rapidly at all positions with those of others who respond more slowly and show less generalization.

At first glance, it might appear that the faster rate of increase in times for the slow responders is contrary to Mednick's (1958) finding of a systematic increase in times only in the case of the fast responders. However, it should be remembered that the two studies are not totally comparable inasmuch as *S*s in the present investigation were required to respond on every trial whereas Mednick's *S*s were not.

It will be noted that there are still a few instances in which an inverse relationship does not obtain. The majority of such exceptions are based on responses to stimuli most distant from the training stimulus and consequently represent times which are based only on a few of the 10 *S*s in each group. In contrast, all or nearly all 10 *S*s are represented in the mean times for positions near the training stimulus. Thus it is entirely possible that an artifact of the combination of data from fast and slow responders is operative not only when all data are grouped together but also when they are broken down as in Table 2; if such an artifact were present, it could account for the inversions. In line with this, it will be noted that the relationship is improved somewhat when the times based only on 1 or 2 *S*s (indicated in Table 2 by parentheses) are excluded from consideration. In general, the fewer the *S*s upon which the mean times are based, the more frequent are the inversions.

The finding that regrouping *S*s into fast and slow responders attenuated the tendency for an inverse relation to break down suggests that the breakdowns reported by Kellogg (1931) and, to some extent, LaBerge (1961) may have been due, at least in part, to an artifact arising from the combination of data from fast and slow *S*s. The fact that Buss (1950) found no evidence of a breakdown even though he did not separate fast and slow *S*s is not particularly suprising inasmuch as he used only one test stimulus between the two training stimuli. Thus it is entirely possible that he did not have enough test stimuli for the breakdown to appear.

Verbal solution.—Each *S* was asked at the end of the experiment to indicate verbally which response he felt was correct for each of the stimulus positions used in testing. The results of these "verbal solutions" are summarized in Table 3. The first column lists the number of *S*s who reported that the solution was to give R_1 to all stimuli on the S_1 side of center, and R_2 to the opposite side. These *S*s felt that no stimuli were exactly in the center. The second column indicates those who reported that the task was to make R_1 to the center and all stimuli to the S_1 side of center, and to make R_2 to the oppostie side. The frequencies of responses which did not fit either of these categories (e.g., R_1-R_2-R_1, R_1-R_2-R_2, etc.) are presented in the third column.

TABLE 3

VERBAL SOLUTION AS A FUNCTION OF REINFORCEMENT CONDITIONS DURING TRAINING

Group	Pattern of Response		
	R_1 - - - R_2	R_1–R_1–R_2	Other
100-0	16	3	1
100-20	9	6	5
100-40	2	9	9
100-60	2	14	4

The more frequently R_1 was correct to S_2, the more frequently *S*s said R_1 was correct for the center position. Since the center position was never presented during training, these data indicate that *S*'s verbal responses display generalization-like effects in much the same way as their motor responses. However, there is at least one important difference. The *S*'s motor responses showed a gradual decline in frequency as the test stimulus became more removed from that used during training. Moreover, this gradual decline was not an artifact of grouped data inasmuch as individual *S*s displayed the same type of gradient. The individual frequency gradients were very similar to those reported by LaBerge (1961).

In contrast to the motor data, the verbal responses were all-or-none in character; i.e., *S* reported either that R_1 was correct all of the time for a particular stimulus, or that R_2 was correct all of the time. No *S* indicated a stimulus position at which he felt R_1 was correct on a portion of the trials and R_2 the remainder of the time.

REFERENCES

BERLYNE, D. E. Conflict and choice time. *Brit. J. Psychol.*, 1957, **48**, 106–118.

BRAINARD, R. W., IRBY, T. S., FITTS, P. M., & ALLUISI, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

BROWN, J. S., BILODEAU, E. A., & BARON, M. R. Bidirectional gradient in the strength of a generalized voluntary response to stimuli on a visual space dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.

BUSS, A. H. A study of concept formation as a function of reinforcement and stimulus generalization. *J. exp. Psychol.*, 1950, **40**, 494–503.

Cartwright, D. Relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174–196.

Crossman, E. R. F. W. Entropy and choice time: The effect of frequency unbalance on choice-response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

Crossman, E. R. F. W. The measurement of discriminability. *Quart. J. exp. Psychol.*, 1955, **7**, 176–195.

Crossman, E. R. F. W., & Szafran, J. Changes with age in the speed of information intake and discrimination. *Experientia*, 1956, Suppl. No. 4, 128–135.

Festinger, L. Studies of decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *J. exp. Psychol.*, 1943, **32**, 291–306.

Fitts, P. M., Peterson, J. R., & Wolpe, G. Cognitive aspects of information processing: II. Adjustments to stimulus redundancy. *J. exp. Psychol.*, 1963, **65**, 423–432.

Gibson, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, **24**, 237–253.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

Johnson, D. M. Confidence and speed in the two-category judgment. *Arch. Psychol., N. Y.*, 1939, No. 241.

Kellogg, W. N. The time of judgment in psychometric measures. *Amer. J. Psychol.*, 1931, **43**, 65–86.

LaBerge, D. Generalization gradients in a discrimination situation. *J. exp. Psychol.*, 1961, **62**, 88–94.

Lemmon, V. W. The relation of reaction time to measures of intelligence, memory and learning. *Arch. Psychol., N. Y.*, 1927, No. 94.

Mednick, S. A. Distortions in the gradients of stimulus generalization related to cortical brain damage and schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **51**, 536–542.

Mednick, S. A. Gradients of latency in a generalized voluntary response. *Amer. J. Psychol.*, 1958, **71**, 752–755.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

(Received October 3, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 485–499

FORM RECOGNITION AT BRIEF DURATIONS AS A FUNCTION OF ADAPTING FIELD AND INTERVAL BETWEEN STIMULATIONS[1]

CHARLES W. ERIKSEN AND MELVIN HOFFMAN

University of Illinois

2 experiments explored the questions as to whether an unrecognized stimulation has an effect upon the recognition of a subsequent stimulation and the duration of the time interval between successive presentations of a weak stimulus in order for the recognitions of the stimulus to be independent of one another. By means of a 3-field tachistoscope the recognition for a form presented twice where the interval between presentations was 0, 5, 250, or 450 msec. was compared with recognition for a single exposure of the form at the same duration. 4 conditions involving variation of the brightness of adapting field and of the interval between form presentations were investigated. In Experiment I, 5 practiced Ss made judgments under all conditions. In Experiment II, 12 Ss were assigned, 3 to each condition. Recognizability was found to be a complex function of whether a form was followed or preceded by a flash of light, the lag time between stimulations and the brightness of the adapting field. The results were explained in terms of brightness summation that resulted in reduced figure-ground contrast for the form. Double presentation was slightly superior to single presentations, the amount of gain being a negatively accelerated function of lag time between the 2 presentations.

The research reported below is concerned with two distinct but nevertheless interrelated problems in visual perception. The first pertains to whether or not a low intensity undetected stimulation leaves any residual effects upon the organism as assessed by changes in subsequent behavior or more specifically changes in the probability of recognition of the signal on reoccurrences. The possibility of such an effect is analogous to the concept of incremental strengthening of responses in learning. The second problem concerns the duration of the time interval (lag) between successive presentations of a weak signal in order for the recognitions of the signal to be independent of one another.

Clinical and personality psychologists have been concerned with the first of these problems in attempting to substantiate their concepts of the unconscious. Experiments such as those of Shevrin and Luborsky (1958), Hilgard (1962), Klein, Spence, Holt, and Gourevitch (1958), and Goldstein and Barthol (1960), have attempted to show that stimulation that is unreported by the human *S* is nonetheless capable of changing or influencing subsequent behavior. As has been pointed out elsewhere (Eriksen, 1960, 1962; Johnson & Eriksen, 1961) these experiments in general have been too poorly controlled to shed much light on the basic question. They have, however, been valuable in pointing up gaps in our knowledge concerning basic perceptual functioning.

The threshold or detection function for various signals typically shows a range of signal intensity where *S*'s detection or recognition is above chance accuracy but below 100%. At this signal intensity level does an unrecognized signal leave the perceptual system in the same state as though the signal had not occurred; or are there residual effects in the system that would be revealed in terms of changed probabilities for recognition of a second signal occurring within defined lag times? One possible effect might be that of summation.

---

[1] This research was supported by Mental Health Grant M-1206.

A possible mechanism for such summation to occur is suggested by the recent work of Sperling (1960), and Averbach and Coriell (1961). These investigators have found evidence of a perceptual storage in the visual system that seems to have a duration in the neighborhood of 250 msec. If the lag time between successive stimulations is within this value, it is possible that this raw memory process could be reinforced or strengthened.

The signal detection theory of perceptual judgments (Egan, Greenberg, & Shulman, 1961) is relevant to the second problem of this investigation. This model dispenses with the concept of the threshold and considers detection essentially as a situation where *S* must discriminate the signal from the general noise background that exists within the sensory perceptual system itself. Thus an *S*'s failure to detect a signal is assumed to be due to the masking of the signal at that moment in time by noise within the perceptual system. If such noise exists in the perceptual system, it is assumed to be variable in intensity and must have a distribution in time. By systematically varying the lag interval between two presentations of a signal it should be possible to obtain a rough mapping of the distribution of this noise envelope. There should be some lag interval beyond which the noise magnitude present on the occurrence of the first signal is uncorrelated with the noise level present at the time of the second signal.

The problem of the independence of successive perceptual judgments is not a new one in psychology. Numerous investigators, e.g., Senders and Sowards (1952), Verplanck, Collier, and Cotton (1952), McGill (1957), and Day (1956), have shown that successive or serial perceptual judgments have a high degree of interdependence. However, these studies have not been successful in showing to what degree these interdependencies of successive responses are due to characteristics of the perceptual process itself as opposed to response effects.

METHOD

Experiment I

Subjects.—Five practiced *S*s, one female, were used in Exp. I. Each gave an equal number of judgments under all combinations of experimental treatments and conditions.

Procedure.—The forms used were the capital letters A, T, and U. Preliminary work had shown that these letters had approximately equal recognizability at brief exposure durations and were approximately equal in confusability, one with another. The experimental procedure made use of a three-field tachistoscope, an adapting field, and two stimulus exposure fields. The adapting field was used to vary the pre- and postadaptation brightness and the forms were presented in Fields I and II, the exposure fields. In addition to permitting precise control of the duration of exposures in Fields I and II, the apparatus also allowed precise control of the interval between termination of stimulation in Field I and onset of stimulation in Field II.

Three kinds of stimulus presentation were used: (*a*) a form presented in Field I followed by light alone in Field II; (*b*) a form presented in Field II preceded by light alone in Field I; (*c*) same form presented in Fields I and II. Intervals or lag times between termination of Field I and onset of Field II of 0, 5, 250, and 450 msec. were studied.

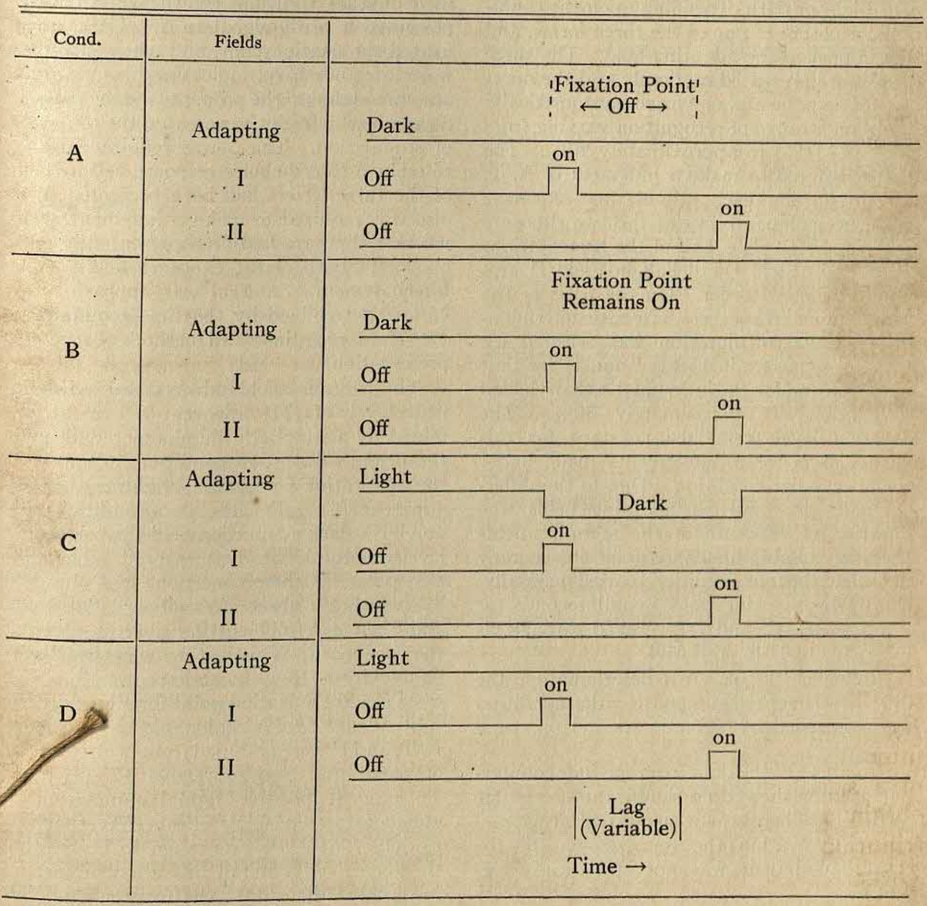
In addition, two other variables were

manipulated. Recognition for single and double presentations at each of the lag times was studied using a dark pre- and postadaptation field with only a faintly glowing fixation point as well as with a relatively bright pre- and postadaptation field. Also the effect of leaving the adapting field on during the stimulus presentations in Fields I and II and during the lag period was compared with the condition where the adapting field went off with the onset of the first field and remained off until the termination of Field II.

Thus, in summary, the three methods of form presentation (Field I, Field II, and double stimulation) at each of the four lag times were studied under four conditions. The conditions were: Cond. A, adaptation field dark with dimly luminous fixation point that terminated with onset of Field I and returned with termination of Field II; Cond. B, adaptation field dark but luminous fixation point remained on during presentation of Fields I and II and during the lag interval between them; Cond. C, adaptation field bright but terminated with onset of Field I and returned with termination of Field II; Cond. D, bright adaptation field that remained on during presentation of Fields I and II and during the lag time between fields. Table I diagrams the events and their time relations for a single trial under each of these four conditions.

TABLE 1

Events Occurring in the Adapting and the Two Stimulus Exposure Fields and Their Relative Chronological Relationships during a Single Trial under Each of the Four Conditions



In Exp. I each *S* sat for 20 40-min. sessions. With the exception of the first, sixth, eleventh, and sixteenth sessions each of the three forms was presented six times under each of the three methods of stimulus presentation and at each of the four lag times. Four successive sessions were devoted to judgments under each of the four conditions and within each session the three methods of presentation and four lag times were counterbalanced. The order in which the different conditions were judged was counterbalanced on the first four *S*s. The fifth *S* followed a different condition order than obtained for the first four *S*s.[2]

The first, sixth, eleventh, and sixteenth sessions initiated a new condition for each *S* and these sessions were devoted to obtaining a function showing *S*'s recognition accuracy as a function of duration of exposure under the conditions of pre- and postadaptation field obtaining for that particular condition. The stimulation sequence was adaptation field (light or dark depending upon the condition), exposure of one of the three forms, and then return to the adapting field. The three forms were presented randomly and the duration of exposure was varied systematically to obtain a range of recognition varying from chance (33⅓%) to approximately 90%. The *S* was required to make a judgment of A, T, or U following each stimulation sequence. To insure comparability of the two exposure fields in the T-scope, half of the presentations occurred in Field I and half in Field II in a counterbalanced order. On the basis of the data obtained from these base rate determinations a level of duration was selected for stimulus exposure in Fields I and II for that condition and for the individual *S* that yielded recognition of approximately 50%. The duration levels remained constant for this particular *S* for all judgments made under that particular condition. Due to the differences in pre- and postadapting fields, the duration of exposure in the stimulus fields necessary to obtain 50% recognition accuracy in each of the four conditions varied markedly. The respective durations in milliseconds for Cond. A, B, C, and D averaged through *S*s were: 14, 14.4, 46, and 61.4.

Following the base rate determination, the four experimental sessions for the condition were preceded by a 7-min. period of visual adaptation to the experimental situation during which time *S* made practice judgments before undertaking the experimental judgments.

The *S* was instructed that he was participating in a study of form perception or recognition. He was told that on each trial one of three letters A, T, or U would be presented and that the letter would occur sometimes in the flash of the first field, sometimes in the flash of the second field, and that on some trials the same letter would occur in both fields. He was further informed that these methods of presentation would vary randomly and also that the interval between flashes would vary. It was stressed that on those trials where a letter appeared in both fields it would always be the same letter in both of the fields.

Judgments were obtained in the following manner. After *E* had arranged for the particular method of stimulus presentation to be used on that trial and set the lag time, he presented a ready signal to *S*. The *S* then fixated the fixation point and when the cross was clear and sharp, indicating his eyes were accommodated to the proper plane, he pressed a thumb switch which presented the sequence of stimulation. The *S* was required always to make a forced-choice response as to which of the three letters had been presented. He also was required to make a judgment as to whether the form had been presented in only one or in both of the exposure fields. The interjudgmental interval was approximately 15 sec. determined by the time required for *E* to change lag times and methods of stimulus presentation.

The exposure fields always had a brightness level of .201 apparent ft-c as determined by a MacBeth illuminometer. Under the bright adaptation conditions (C and D) the adaptation field had a brightness of .582 apparent ft-c. For the two conditions involving a dark adaptation field the brightness level was too low to be measured. The fixation point for these conditions was obtained by painting a sheet of cardboard flat black and then carefully cutting a cross through the cardboard. This in turn was then backlighted by a 10-w. incandescent bulb. The cross used as a fixation point for all conditions subtended 12′ of visual angle. The letters A, T, and U were obtained from printed sheets of paper and also subtended 12′ of angle. To prevent possible retinal summation or metacontrast masking effects (Averbach & Coriell, 1961) care was taken to insure that the forms did not fall on retinal areas immediately adjacent to the fixation stimulus.

---

[2] Analysis of the data from the four counterbalanced *S*s showed no significant effects due to order. There was a suggestion of a practice effect within the four sessions of the conditions but it was not statistically significant.

This was arranged by presenting the form on the points of an imaginary square of 1.25° of angle centered on the fixation point. Thus a form stimulus could occur in any one of these four possible positions determined on a random basis with the restriction that for double stimulation the same position would not occur for both presentations.

Apparatus.—The three-field tachistoscope was built to special order by Merle Ridgely & Company. Two 4-w. fluorescent tubes were used in each of the exposure fields which provided square onset of illumination although a somewhat tapered decay. During experimentation the timing section of the T-scope was periodically checked for calibration by means of photocells and a Hunter klockounter. Duration settings throughout the range employed in this experiment were found to be exceedingly stable and to have less than a 5% maximum error. Due to limitations in the timing equipment the return to the adapting field illumination following termination of the flash in Field II was delayed in the order of 250–300 msec. The three forms were each mounted on plastic cards off-white in color. The size of the exposure fields and the adaptation field was a square subtending 7.6° of angle on each dimension.

Experiment II

Subjects.—Twelve undergraduate students at the University of Illinois, five females, were used in this study. All were experimentally naive with respect to perceptual experiments but before undertaking the present experiment they were given two practice sessions during which they made judgments of tachistoscopically-exposed stimuli.

Procedure.—The procedure and method in this experiment were identical to that of Exp. I except for the statistical design. Here three different *S*s were assigned to each of the four conditions. Within a condition each of the three *S*s had 24 presentations of each of the three forms under each of the three methods of presentation for each of the four lag times. The durations of exposures in milliseconds in the stimulus fields necessary to obtain 50% recognition under Cond. A, B, C, and D were, respectively: 18.8, 18.3, 40, and 76.

RESULTS

In both Exp. I and II the number of correct recognitions was summed through the three forms and through sessions. In Exp. I these recognition scores were analyzed by a four-way classification analysis of variance (method of presentation, lag time, condition, and *S*s). The summary of this analysis is given in Table 2. For Exp. II, where a different statistical design was employed, the recognition scores were analyzed by a modified four-way classification analysis of variance. The summary of this analysis is contained in Table 3.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF RECOGNITION: EXP I

Source	*MS*	*F*
Lag times (L)	117.2	20.21*
Fields (F)	130.2	22.44*
Conditions (C)	48.7	8.40*
Individuals (I)	63.1	10.88*
L × F	8.1	1.40
L × C	21.4	3.69*
L × I	9.5	1.64
F × C	59.3	10.22*
F × I	6.3	1.09
C × I	18.1	3.12*
L × F × C	9.1	1.57
L × F × I	4.2	<1.00
L × C × I	9.2	1.59
F × C × I	6.6	1.14
Error	5.8	

$*p < .05.$

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF RECOGNITIONS: EXP. II

Source	*MS*	*F*
Lag times (L)	113.4	9.56*
Fields (F)	100.5	7.89*
Conditions (C)	64.9	1.79
Individuals (I)	36.2	7.67*
L × F	9.6	2.03
L × C	15.9	1.34
F × C	30.6	2.40
L × F × C	7.3	1.55
Error 1	11.86	
Error 2	12.74	
Error 3	4.72	

$*p < .05.$

In both experiments the main effects of lag times and method of form presentation (fields) are significant ($p < .05$). Further, conditions is a significant effect in Exp. I although failing to achieve significance in Exp. II. This discrepancy was anticipated due to the difference in statistical design between the two experiments. In Exp. I the same five *S*s gave data under all four conditions whereas three different *S*s were employed for each of the four conditions in Exp. II, thus making this second experiment less sensitive for testing condition effects.[3]

In the first experiment the second-order interactions of Lags × Condition and Method of Stimulation (fields) × Condition are significant. In Exp. II neither of these interactions reaches significance although, as will be seen in Fig. 1, the relationships between these variables are nearly identical in both experiments. There is a suggestion ($p < .10$) in both experiments of an interaction of Lags × Method of Stimulation × Conditions.

The nature of the results can be seen more clearly in Fig. 1. Here, number of correct recognitions has been plotted as a function of method of stimulation and lag time by condition for each of the two experiments. Percent correct recognitions obtained at the base rate level are plotted for purposes of comparison. The suggestion of a triple interaction in statistical analyses taken in conjunction with the very marked similarity of the functions from the two experiments indicates that interpretation of the effects of method of stimulation and of lag time is best made separately for the four conditions.

The effect of lag times and method of stimulation appears quite similar for both Cond. A and B. Under both conditions and in both experiments there is a definite tendency for recognition to improve irrespective of the method of stimulation as lag time increases from 0 to 450 msec. There is also a slight tendency for the double method of presentation to be slightly superior to that obtained for a single presentation in either field. However this superiority is not marked. In Cond. C there is again a marked tendency for increasing lag times to result in increased recognition of forms presented twice (double stimulation) and with forms presented in Field II, but here there is a discrepancy for Field I. Increasing lag time does not improve recognition of forms appearing in Field I.

Condition D is the most discrepant of the four conditions. Here none of the three methods of stimulation show increasing recognition as a function of increasing lag time. However for both Cond. C and D there is again a slight tendency for the double stimulation to be slightly superior.

In addition to requiring a forced choice as to which of the three forms had been presented on each trial *S*s were required to make a judgment as to whether the presentation had been a single form stimulation or a double presentation of the form, i.e., once in each of the two exposure fields. To determine whether *S*s were capable of making reliable discriminations as to single and double stimulation the number of single judgments to each of the three methods of presentation was averaged through forms and

---

[3] The significant main effect due to *S*s in Exp. I and the *S* × Condition interaction can be attributed to our inability to set the duration of the exposure fields at exactly the same identifiability level for each *S* and for same *S* under all four conditions. With the exception of a significant *S* × Condition interaction the remaining *S* interactions are all nonsignificant indicating that the experimental effects had a consistent result across all five *S*s.

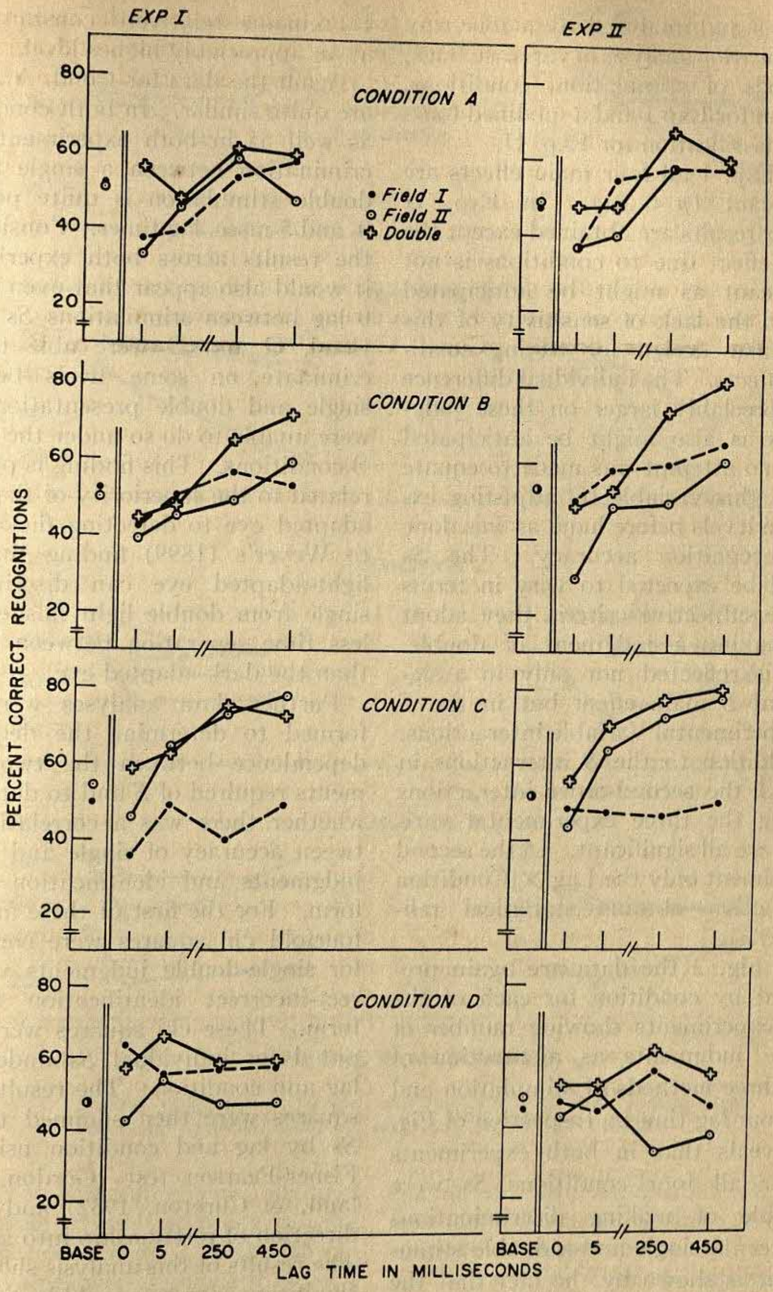


FIG. 1. Percent correct form recognitions as a function of method of presentations, time lags between fields, and adapting brightness condition.

sessions and analyzed by a four-way classification analysis of variance (lags, methods of stimulation, conditions, and *S*s) for Exp I and a modified four-way classification for Exp. II.

In Exp. I all four main effects are significant ($p < .01$). In Exp. II similar results are obtained except the main effect due to conditions is not significant as might be anticipated due to the lack of sensitivity of this particular design for testing condition effects. The individual difference is appreciably larger on these judgments as also might be anticipated since no attempt was made to equate *S*s on this variable by adjusting exposure levels before hand as was done for recognition accuracy. The *S*s would be expected to vary in terms of the subjective criteria they adopt for making a judgment of double. This is reflected not only in a significant *S* main effect but in the *S* × Experimental Variable interactions. In addition to the *S* interactions in Exp. I the second-order interactions among the three experimental variables are all significant. In the second experiment only the Lag × Condition interaction obtains statistical reliability.

In Fig. 2 the data are again presented by condition for each of the two experiments showing number of single judgments as a function of the three methods of stimulation and the four lag times. Inspection of Fig. 2 reveals that in both experiments under all four conditions *S*s were capable of making discriminations between a single and a double stimulation as shown by the fact that the number of single judgments to the double stimulation decreases with increasing lag times whereas the number of single judgments to a single stimulation in either Field I or Field II remains relatively constant and at an appreciably higher level.

Again the data for Cond. A and B are quite similar. In both conditions as well as in both experiments discrimination between a single and a double stimulation is quite poor at 0- and 5-msec. lag times. Considering the results across both experiments it would also appear that even at the 0 lag between stimulations *S*s under Cond. C and D were able to discriminate on some trials between single and double presentations but were unable to do so under the A and B conditions. This finding is perhaps related to the superiority of the light-adapted eye to detecting flicker and to Weyer's (1899) finding that the light-adapted eye can discriminate single from double light flashes with less time separation between flashes than the dark-adapted eye.

Further data analyses were performed to determine the degree of dependence between the two judgments required of *S* and to determine whether there was a correlation between accuracy of single and double judgments and identification of the form. For the first of these analyses fourfold chi squares were computed for single-double judgments vs. correct-incorrect identification of the form. These chi squares were computed for individual *S*s under each lag and condition. The resulting chi squares were then summed through *S*s by lag and condition using the Fisher-Pearson test (Gordon, Loveland, & Cureton, 1952) and taking direction of relationship into account. The results of this analysis showed no significant nor appreciable tendency for *S*'s same-difference judgments to be correlated with accuracy of form identification.

A similar chi square analysis was undertaken to determine whether

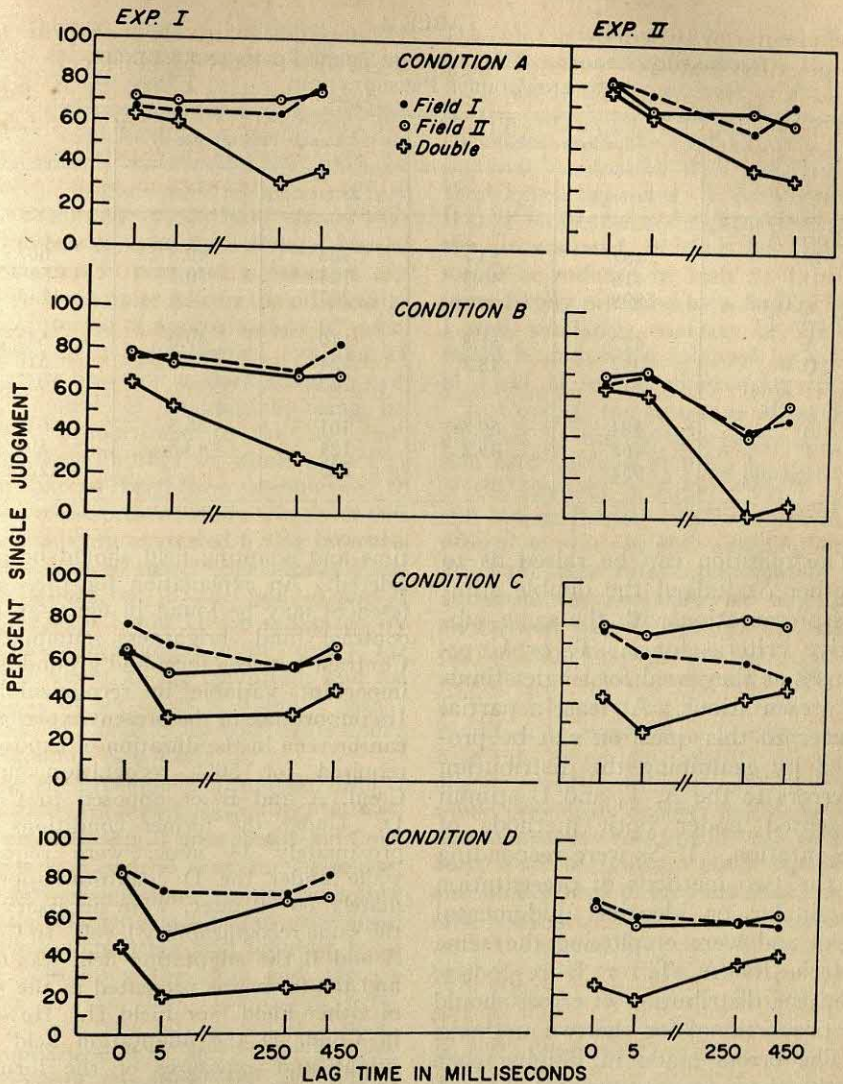


FIG. 2. Percent judgments of "single" as a function of method of stimulation, time lags between fields, and adapting brightness condition.

accuracy in same-difference judgment was correlated with accuracy in form identification. Here 2 of the 16 chi square values were significant beyond the .05 level for Exp. I and 3 for Exp. II. Considering that the significant chi squares did not cluster in particular lags and conditions within and across experiments and also considering the number of repeated comparisons that were made, these five significant relationships would appear to be within chance expectancy. Accordingly we can conclude there is no evidence that the accuracy of a single-double judgment to a presentation is related to *S*'s recognition accuracy on that presentation.

TABLE 4

RECOGNITION ERRORS IN EXP. I TO THE THREE FORMS UNDER SINGLE AND DOUBLE PRESENTATIONS

Responses	Single Presentation		Double Presentation		Total Errors
	Errors	%	Errors	%	
Form A					
T	482	53.7	123	49.8	605
U	416	46.3	124	50.2	540
Total	898		247		1145
Form T					
A	422	51.8	119	49.0	541
U	392	48.2	124	51.0	516
Total	814		243		1057
Form U					
A	484	50.8	161	56.5	645
T	468	49.2	124	43.5	592
Total	952		285		1237

The question can be raised as to whether *S*s judged the double stimulus presentation with the same subjective criteria for A, T, or U responses as they used for single stimulus presentations. At least a partial answer to this question can be provided by examining the distribution of errors to the A, T, and U stimuli presented under both methods of presentation. If *S*s were responding to the two methods of presentation as similar or identical judgmental tasks and were employing the same criteria for A, T, or U responses, then the distribution of errors should be proportional for the two methods.

The errors made in judging each of the three forms, summed through fields, lags, and conditions for Exp. I *S*s, are shown in Table 4. Inspection of these data shows that incorrect recognitions or errors are distributed virtually the same for single and double stimulations.

DISCUSSION

Before considering the effects of double stimulation, the differences in form recognition for stimuli presented only in Field I or Field II as a function of lag time and adapting field should be considered. An explanation for these differences may be found in figure-ground contrast and brightness summation. Contrast between form and ground is an important variable in recognizability. Its importance in the present experiment can be seen in the duration of exposures required for 50% recognition under Cond. A and B as opposed to Cond. D. Under the former conditions approximately 14 msec. were required while under the D condition approximately 60–70 msec. were necessary for the same recognizability level. In Cond. A and B the adaptation field was dark and the form was presented in the flash of either Field I or Field II. However in Cond. D the adaptation field was bright and exposures of the form in Field I or II were superimposed upon the illumination of the adapting field. Thus the illumination of the adapting field added equal increments to the light and dark sections of the stimulus which reduced figure-ground contrast. The differences in exposure required for a given level of recognizability between these conditions cannot be explained in terms of light adaptation since Cond. C employed the same adapting field as the D condition but here the stimulus exposures were not superimposed upon the illumination of the adapting field.

Nonetheless Cond. C required appreciably shorter durations for 50% recognizability (40 msec.).

The differences in form recognizability can be explained in terms of contrast differences if we consider the phenomenon of brightness summation. Clark and Blackwell (1959) have shown that two pulses of light show at least partial summation in terms of a detection criterion if one pulse follows the other within an interval of approximately 65 msec. or less for a light-adapted eye and 85 msec. or less for a dark-adapted eye, the amount of summation being inversely proportional to the lag time. From their findings we would expect at least partial brightness summation to occur in the present study whenever two light fields are separated by lag intervals of less than 80 msec. Specifically these cases would be: (*a*) in Cond. A and B when Fields I and II are separated by lags of 0 and 5 msec.; (*b*) in Cond. C for Field I under all lag conditions and for Field II at lag times of 0 and 5 msec.; and (*c*) in Cond. D for both fields under all lag times (see Table 1).

The effectiveness of a contrast brightness-summation explanation can be seen if the data in Fig. 1 are examined. Consider first Cond. A and B. Here the base rate recognition levels for each field were determined separately by presenting the dark adaptation field followed by a flash containing the form and then a return to the dark adaptation field. Using these base rate values as a point of comparison it is noted that in both experiments recognizability in Field I and in Field II tends to be depressed below the base value at lag times of 0 and 5 msec. This is what would be expected since the experimental stimulus presentations at these two lag times should result in brightness summation between Field I and Field II with a reduction in figure-ground contrast due to the fact that whichever of the two fields contains the stimulus, the flash of the other field, through this brightness summation, adds equal increments to both the light and dark areas of the following or preceding stimulus-containing field. However at lag times of 250 and 450 msec. recognizability in the two fields has improved at least to the point of the base rate levels. This is again consistent with the explanation assuming that summation does not occur at these longer lag times. It is to be noted that a bidirectional brightness summation is assumed so that the contrast would be reduced in Field II if it was immediately preceded by a flash in Field I and similarly, contrast in Field I would be reduced if followed by a flash in Field II at the shorter lag times.[4]

In Cond. C the base rate values were obtained by the use of a bright adaptation field followed at 0 lag by exposure of the stimulus in either Field I or II and approximately 250 msec. later a return to the bright adapting field. Thus the base rate values contain the circumstances necessary for brightness summation, namely between the adapting field and the exposure field. Recognizability of forms presented in Field I at 0 and 5 msec. lag should under this condition be expected to be close to the base rate or perhaps slightly depressed due to the flash in the second field which would provide an opportunity for brightness summation to occur at both ends of Field I presentation. Field II would also be expected to give essentially base rate level at 0 and 5 msec. lag since again the summation would be occurring from the flash in Field I that preceded a presentation of a form in Field II. However at the longer lag intervals recognizability for Field II presentations should improve relative to the base rate level whereas for Field I they should tend to remain constant since Field I is always immediately preceded by the bright adapting field.

---

[4] Since the fluorescent tubes in the T-scope lacked an immediate decay due to some persistence of the phosphor coating, the brightness summation may have been in part due to the physical summation of light. However, an appreciable part of the summation must be attributed to the characteristics of the visual system since 40% of the brightness of the tubes decayed in a fraction of a millisecond and the remaining decay was no longer measurable with a fast photocell and oscilloscope after 3 msec.

In Cond. D the results again are quite consistent with our explanation. Here the base rate level was obtained by superimposing a stimulus exposure upon the bright illumination of the adapting field. Irrespective of lag times in the experimental judgments, conditions existed for brightness summation in whichever field the form was presented. Even at 450 msec. lag, presentation of a form in Field II is equivalent to the presentation of a form in Field I under Cond. C and the base rate situation for Cond. D. As a consequence, no improvement over the base rate levels in recognizability would be expected as a function of lag time.

The results in Fig. 1 fit quite well with the contrast brightness-summation explanation. Minor discrepancies such as a tendency for recognizability to rise somewhat above the base rate level at longer lag times in Cond. A and B and not to show appreciable reduction below base levels at the 0- and 5-msec. lag times would seem to be attributable to practice effects. The experimental judgments were always made following the base rate session so that *S*s had opportunity to become more proficient at perceptual recognition under the particular viewing conditions. Also it is to be remembered that due to limitations in the timing equipment the return of the adapting field in Cond. C following the flash in Field II was delayed in the order of 250–300 msec. This time delay is consistent with the findings that there would be no brightness summation effects due to the return of the adapting field for stimulation presented in Field II at the longer lag times.

Another possible explanation for the above discussed effects might be found in terms of dark adaptation which would be expected to yield better recognition. However a dark adaptation explanation is contradicted by several findings in the data. Dark adaptation would have to work in reverse in order to explain the gain in form recognizability in Field I stimulations under Cond. A and B. Also in the A, B, and C conditions 5 msec. of lag or dark time gives an appreciable increase to recognizability of forms in Field II and it seems questionable whether enough dark adaptation could occur in 5 msec. to account for this large increase in recognizability.

Another possible explanation might be advanced in terms of *S*s blinking to the flashing light in Field I. If the flash of light in Field I caused occasional blinks in *S*s, this would be expected to effect the recognizability of forms presented in Field II when the lag time corresponded with the latency of the blink. However an explanation along these lines would be unable to explain why there is a gain in recognizability of forms presented in Field I under the A and B conditions and also the latency for the blink would have to be in the order of approximately 14 msec. in these conditions and approximately 50 msec. in Cond. C to appreciably effect the results.

Another finding of interest in the Fig. 1 data is the apparent ability of an event to act backward in time and effect the recognizability for forms that preceded. In Cond. A and B the flash of light from Field II following within 5 msec. of a form presented in Field I impairs recognition of the form. Since this form had been presented for approximately 14 msec. in Field I with full contrast before any brightness summation could be present, this would imply that the 14-msec. period was insufficient for consolidation of perception to occur. It further implies that the recognition level achieved during the base rate level assessments under these conditions was achieved not in the approximately 14 msec. exposure of the form but actually occupied part of the dark period following the exposure. (Concern with this problem dates back to Wundt's laboratory and Sperling, 1960, has provided a recent review of this older literature.) This effect is quite similar to the metacontrast phenomena that have been reported by various investigators (Alpern, 1952; Toch, 1956; Werner, 1935) and some of the metacontrast phenomena might well reduce to the

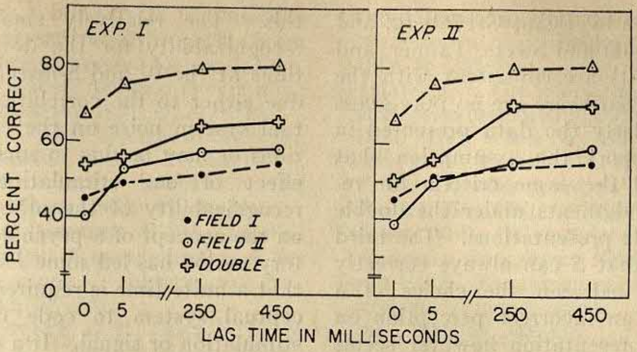


FIG. 3. Percent correct form recognitions for single and for double stimulations.

effects of figure-ground contrast and brightness summation that seem to be involved in our results.

The recognizability of forms presented with the double stimulation as opposed to a single stimulation in either field was found to be slightly but consistently superior at most lag times. This can be seen most readily in Fig. 3 where we have presented separately for the two experiments the percent correct recognitions for forms presented in Field I, in Field II, and for double presentation as a function of lag time. The data points have been summed through forms, *S*s, and conditions in order to obtain the most stable values. The interpretation of these gains in recognizability with double presentation must consider that *S* had two opportunities to correctly perceive the form. If we wish to raise the question whether an incorrectly recognized stimulus has some residual effects in the nervous system as manifested in the recognition behavior for a subsequent presentation of the stimulus, in other words a summative or perhaps inhibitory effect, then we need to know the level of recognition that would be expected if the two opportunities were independent perceptual events. Similarly the question of how long a time duration is necessary between two stimulations in order for error in the perceptual system to be uncorrelated requires an estimate of what the recognizability level should be for a double stimulus presentation if the two events are indeed independent. The lag time at which the empirical level reaches the level for independence would then indicate the lag time necessary for noncorrelated error.

A computation of the recognition level to be expected under the case of independent perceptual events poses a number of difficulties. Clark and Blackwell (1959) employed a standard formula for statistical independence which applied to the present situation would yield:

$$P_D = P_{F1} + P_{F2} - (P_{F1} \cdot P_{F2})$$

where P_D is the proportion of correct recognitions for the double presentations and P_{F1} and P_{F2} are the observed proportion of correct recognitions for stimulation in Fields I and II, respectively. To use this formula for the present data would require making three important assumptions concerning the perceptual recognition process. First it would be necessary to assume that there are no guess trials; second that *S* uses the same criteria for double as opposed to single stimulus presentations; and third that when *S* has an incorrect and a correct perception as a result of the double stimulation and is allowed only one response he always decides in favor of the correct perception.

The first two assumptions are not too implausable. The concepts from the theory of signal detection as applied to psychophysics by Egan, Greenberg, and Shulman (1961) and Swets, Tanner, and Birdsall (1961) are consistent with the assumption that there are no pure guess trials. Similarly the data presented in Table 4 support the assumption that *S*s employed the same criteria for responses or judgments under the double and the single presentations. The third assumption that *S* can always correctly discriminate between the choice of a correct and an incorrect perception on the double presentation however seems much less tenable.[5]

In Fig. 3 the upper curve represents the level of recognizability that would be predicted for double presentations employing the formula for statistical independence. While the use of this formula for predicting the case of independence is greatly restricted by the third assumption discussed above, the curve nonetheless provides an anchor point for evaluating the empirically determined level of performance.

The curves in Fig. 3 do show that double stimulation results in proportionately greater recognizability relative to single stimulation in either field as lag time increases. There is a suggestion that an asymptote is reached or approached at the 250–450 msec. level but further experimentation with longer lag times would be needed to substantiate this. The relatively smaller gain in recognizability for the double stimulations at the 0- and 5-msec. lag may be due either to the correlation of perceptual system noise on the two presentations or may be due to some interfering effect of one stimulation upon the recognizability of the other. Research on the concept of a psychological refractory period has led some *E*s to speculate that a finite time is required by the perceptual system to code or process a stimulation or signal. If a second signal should arrive during this processing time it is either degraded or ineffective. If such an effect does exist, it could explain the results of the present experiment but Clark and Blackwell's (1959) data on the detection function for double presentations of flashes of light would suggest that the concept was limited to the recognition as opposed to the detection situation.

---

[5] A model for the case for independent perceptual events could also be obtained using one of the formulas for guessing but such a formula assumes there are two kinds of trials: Those on which *S* perceives the stimulus and those upon which he gives a guess. The tenability of such an assumption is quite dubious in view of the work during the past several years in which the theory of signal detection has been fruitfully applied to threshold detection situations. Unfortunately the theory of signal detectability itself does not provide a ready model for the three-category recognition situation with which we are confronted in the present experiment. Tanner (1956) has made an extension of the theory to signal recognition for the two-category case, but the fruitfulness of this approach for the recognition situation as well as its extension to three or more categories yet remains to be demonstrated.

REFERENCES

Alpern, M. Metacontrast: Historical introduction. *Amer. J. Optom.*, 1952, **29**, 634.

Averbach, E., & Coriell, A. S. Short-term memory in vision. *Bell Sys. tech. J.*, 1961, **40**, 309–328.

Clark, W. C., & Blackwell, H. R. Visual detection thresholds for single and double light pulses in the temporal element contribution hypothesis. *U. Mich. Engng. Res. Inst. Proj. Mich. Rep.*, 1959, No. 2144–343-T.

Day, W. F. Serial non-randomness in auditory differential thresholds as a function of interstimulus interval. *Amer. J. Psychol.*, 1956, **69**, 387–394.

Egan, J. P., Greenberg, G. Z., & Shulman, A. I. Operating characteristics, signal detectability, and the method of free response. *J. Acoust. Soc. Amer.*, 1961, **33**, 993–1007.

Eriksen, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychol. Rev.*, 1960, **67**, 279–300.

Eriksen, C. W. Figments, fantasies, and follies: A search for the subconscious mind. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N. C.: Duke Univer. Press, 1962. Pp. 3–26.

Goldstein, J. J., & Barthol, R. P. Fantasy responses to subliminal stimuli. *J. abnorm. soc. Psychol.*, 1960, **60**, 22–26.

Gordon, M. H., Loveland, E. H., & Cureton, E. E. An extended table of chi square for two degrees of freedom for use in combining probabilities from independent samples. *Psychometrika*, 1952, **17**, 311–316.

Hilgard, E. R. What becomes of the input from the stimulus? In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N. C.: Duke Univer. Press, 1962. Pp. 46–72.

Johnson, H., & Eriksen, C. W. Preconscious perception: A re-examination of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1961, **62**, 497–503.

Klein, G. S., Spence, D. P., Holt, R. R., & Gourevitch, S. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–266.

McGill, W. J. Serial effects in auditory threshold judgments. *J. exp. Psychol.*, 1957, **53**, 297–303.

Senders, V. L., & Sowards, A. Analysis of response sequences in the setting of a psychological experiment. *Amer. J. Psychol.*, 1952, **65**, 358–374.

Shevrin, H., & Luborsky, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.

Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

Swets, J. A., Tanner, W. P., & Birdsall, T. G. Decision processes in perceptions. *Psychol. Rev.*, 1961, **68**, 301–340.

Tanner, W. P., Jr. A theory of recognition. *J. Acoust. Soc. Amer.*, 1956, **28**, 882–888.

Toch, H. H. The perceptual elaborations of stroboscopic presentations. *Amer. J. Psychol.*, 1956, **69**, 345–358.

Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, **44**, 273–282.

Werner, H. Studies on contour: I. Qualitative analyses. *Amer. J. Psychol.*, 1935, **47**, 40–64.

Weyer, E. M. Die Zeitschwellen gleichartiger und disparater Sinneseindrucke. *Phil. Stud.*, 1899, **15**, 68–138.

(Received December 7, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 500–505

DOES VISUAL AFTEREFFECT OF MOTION ADD ALGEBRAICALLY TO OBJECTIVE MOTION OF THE TEST STIMULUS?

THOMAS R. SCOTT, AUGUSTUS E. JORDAN, AND DONNIE A. POWELL

Veterans Administration Hospital, Columbia, South Carolina

Carlson (1962) found that additivity occurs only when eliciting motions and test motions are in the same direction. This result seemed to conflict with Gibson's general theory of adaptation. In the present study 6 Ss viewed a moving stimulus which was alternated with an electronically produced test stimulus. Comparisons of the velocity of the test stimulus with a previously seen standard were made. The results of this experiment were contrary to Carlson's results and upheld Gibson's theory. In a 2nd experiment, aftereffect measurements were made in a monkey. The data obtained from this animal also clearly supported Gibson's theory. Possible reasons for Carlson's failure to demonstrate algebraic additivity were discussed. It was concluded that Carlson's data provide no serious refutation of Gibson's theory.

Visual aftereffects of motion have been reported and well documented for many years. Following fixation of a moving stimulus pattern (eliciting stimulus), the aftereffect is usually demonstrated in one of three ways: (*a*) a stationary test pattern is presented and appears to move in the opposite direction to the eliciting motion; (*b*) a test pattern moving slowly in the same direction as the eliciting motion appears stationary; or (*c*) the eliciting stimulus pattern continues and an identical moving test pattern, presented in a different part of the visual field, appears to move faster than the eliciting stimulus.

Gibson (1959) has suggested that following the continuous presentation of an unchanging stimulus there is a shift in the correspondence between a dimension of stimulation and a variable of experience. This implies that adaptation to a stimulus not only reduces the difference between the originally perceived value of the stimulus and zero, but also shifts all values on the other side of zero in the same algebraic direction. For example, if an *S* judges the temperature of three pails of water to be cold, neutral, and hot and then undergoes a period of adaptation to very hot water, his judgments might be shifted, respectively, to very cold, cool, and warm. In this example, not only was the original perception of "hot" shifted toward zero (neutral), but the water originally judged "cold" was now perceived as "very cold." The shift from cold to very cold was in the same algebraic direction as the shift from hot to warm but was on the other side of zero from the adapting stimulus. Applying this principle to visual aftereffect of motion, one would predict that a test object moving in the opposite direction to the eliciting stimulus (and thus in the same direction as the aftereffect) would have a perceived velocity greater than its objective velocity.

Carlson (1962) tested the effect of adaptation to an orbitally rotating pattern on a subsequently presented moving test stimulus. He found that when eliciting and test motions were in the same direction, the results were compatible with Gibson's theory of perceptual adaptation in that the test stimulus was seen as moving slower than its objective velocity. However, when adaptation and test motions were in opposite directions, the expected perceived speeding of the test stimulus was not found. He suggested that Gibson's theory may not apply to visually perceived velocity.

The present article contains two experiments: one which measures the influence of motion aftereffect on the perception of a stimulus object moving in the opposite direction to the eliciting motion, and another in which both directions of motion of the test stimulus relative to the eliciting stimulus are used. As will be seen in the following description of the methods used, the measures do not depend upon subjective awareness of the occurrence of aftereffect. It was predicted that the aftereffect would add algebraically to the objective velocity of the test stimulus. Results of this nature would be consistent with Gibson's theory of perceptual adaptation.

EXPERIMENT I

Method

Apparatus.—The apparatus consisted of a device for alternate presentation of a rotating spiral (eliciting stimulus) and a circular oscilloscopic figure (test stimulus). A black four-turn single spiral, drawn counterclockwise outward on a white disc 8 in. in diameter, was rotated at 160 rpm counterclockwise by a stroboscopically controlled motor. The spiral was set in one arm of a mirror tachistoscope and was illuminated by neon lights. The face of an oscilloscope was mounted in the other arm of the tachistoscope and was connected so as to project a 3-in. circular stimulus onto the mirror. This circle could be caused to expand or contract at selected rates. Only motionless or expanding circles were used in this first experiment. The rates used were 0, 2, 4, and 6% per second based on the initial diameter of 3 in.

A timing mechanism was used to present the spiral for a 10-sec. period and then to replace it immediately with the circle for a 2-sec. period. The spiral could be presented rotating or motionless. For more detail as to the apparatus see Scott, Bragg, and Smarr (1963) and Scott and Medlin (1962).[1]

Subjects.—The *S*s consisted of five members of the psychology staff at the VA Hospital in Columbia, S. C., and one student from an introductory psychology class at the University of South Carolina.

Procedure.—The *S*s were seated in a dark room in front of the tachistoscope so that the stimuli appeared at an optical distance of 5 ft. They were instructed to fixate the center of the spiral when it was present.

An initial training period was begun with a 2-sec. presentation of a circle expanding at 4% per second. The *S* was told that this was to be a standard rate of expansion to which he was to compare other rates. After several presentations of the standard, other rates of expansion (2 and 6%) were alternated with it and *S* was asked to judge whether they were expanding faster or slower than the standard rate. All *S*s made all comparisons correctly during the training period.

During the actual testing procedure each trial consisted of a presentation of the standard circle for 2 sec., followed by a 10-sec. presentation of the spiral, followed by a 2-sec. presentation of a comparison circle. On odd numbered trials the spiral was stationary and on even numbered trials it was rotating counterclockwise at 160 rpm. Between trials a stationary spiral was presented for 10 sec. The *S* was asked to judge whether each comparison circle was expanding faster or slower than the preceding standard. Equal judgments were not allowed. Comparison circles which were presented after stationary spirals had speeds of 2, 4, and 6% expansion. Comparison circles following rotating spirals had speeds of 0, 2, and 4% per second expansion. These two ranges were chosen because in preliminary work it was found that speeds beyond these ranges were called "faster" or "slower" 100% of the time. The comparison speeds were presented in a random order, 10 times each, under both the rotation and stationary spiral conditions, thus making a total of 60 trials. The nature of the task insured that *S* could not tell whether any perceived motions were real, illusory, or some combination of the two.

---

[1] A 3-page description and one circuit diagram have been deposited with University Microfilms, Inc., Ann Arbor, Michigan. Order Special Film S-294, remitting $2.75 for microfilm or positive copies.

TABLE 1

Velocity of Test Stimulus (in Percent per Second) When Judged Equal to Standard (4% per Second)

Ss	After Viewing Stationary Spiral	After Viewing Rotating Spiral	Difference
1	3.44	3.64	−.20
2	4.60	3.20	1.40
3	3.92	3.00	.92
4	4.25	2.00	2.25
5	4.00	1.60	2.40
6	4.40	3.18	1.22
M	4.10	2.77	1.33

Results

Each *S*'s point of subjective equality, both after rotating and after stationary spirals, was computed using Spearman's (1908) method for obtaining the arithmetic mean of an ogive. Table 1 shows the results of these computations.

The result, $t\ (5) = 3.43$, $p < .01$, upholds the prediction that motion aftereffect adds to the perceived velocity of a test stimulus moving in the opposite direction to the eliciting motion. The average circle speed, following fixation of the stationary spiral that was judged as equal to the standard circle speed of 4% per second, was 4.1% per second. The comparable average following the rotating spiral condition was 2.8% per second. This means that a circle actually moving at this objectively slower rate of speed was perceived as moving at about 4% per second under the rotating spiral condition. The only apparent explanation for this is that the aftereffect added to the perceived velocity of the test circle.

Experiment II

A further and perhaps more general test of Gibson's theory was provided by the following data obtained from a study of visual motion aftereffect in a monkey.

Method

Subject.—The *S* was a 3-yr.-old male rhesus monkey who had been previously trained to perform the discrimination task involved in this experiment.

Procedure.—The monkey, seated in a restraining chair, looked into a three-field mirror tachistoscope. He fixated a point of light at the center of a four-turn spiral 8 in. in diameter, seen at an optical distance of 5 ft. The spiral could be presented stationary, or rotating clockwise or counterclockwise at 180 rpm. Clockwise rotation would be expected to produce an aftereffect of contraction, counterclockwise rotation would produce perceived expansion. At the end of a 10-sec. period of steady fixation, the spiral and fixation point vanished and were instantly replaced by a 3-in. diameter circle produced on the face of an oscilloscope. The circle could expand or contract at preselected linear rates, or it could remain motionless. The rates used were 2, 4, 6, and 10% per second expansion and contraction. In front of the monkey at hand level was a panel containing a food cup in its center and a lever on each side. To make a "correct" response, the monkey pressed the right hand lever when an expanding circle appeared and the left hand lever for a contracting circle. A whole diet food pellet was released into the food cup for every second "correct" response. A lever response turned off the circle and caused the spiral and fixation point to reappear for the next trial. If the monkey did not respond with a lever press, the circle vanished after 4 sec. and another trial was begun. Either an "incorrect" response or a lever press during the spiral fixation resulted in a mild electric shock. A lever press made during the presentation of a motionless circle resulted in neither shock nor food.

The animal was tested for 12 days. On odd numbered days the spiral was stationary and only 2% per second circle speeds were presented to keep up the desired sharpness in discrimination. On each even numbered day the animal performed under three conditions: one with the spiral stationary, a clockwise rotation condition, and a counterclockwise condition. These conditions were presented in the six possible orders on the 6 even numbered days. The order of the circle speed settings always followed the same counterbalanced program for each condition. During the conditions involving rotation of the spiral, shock was not administered for responses made to 2 or 4% circle rates since this would have constituted negative reinforcement for responses indicative of aftereffect. This change in reinforcement schedule for the rotation conditions was trivial since it resulted in no appreciable change in the proportion of shocked responses. In preliminary studies even complete elimination of shock was found to have no detectable effect on the animal's performance during runs of similar length to those of the experiment proper.

Results

Figure 1 is based upon the data obtained on the even numbered days and shows the results of 1,442 presentations of the circle in response to which 1,406 lever responses were made. Almost all of the 36 failures to respond occurred on difficult discriminations. The ordinate shows the proportion of responses made with the left lever. The abscissa represents rate of circle size change in percent per second, contraction being indicated by the negative sign. The open circles on the graph show the results from the stationary spiral condition. It can be seen that when the spiral was not turning the animal seldom made an error in judging circles whose rate of expansion or contraction was 6% per second or more. The nearly equal left and right lever responses to the motionless circle indicate that little, if any, lever preference was present. A circle speed to which the monkey responds equally often with the left and right levers is used as a measure of his point of perceived zero motion. This point on the ogive of Fig. 1 can be estimated by interpolation. The point of perceived zero motion when the spiral was not rotat-

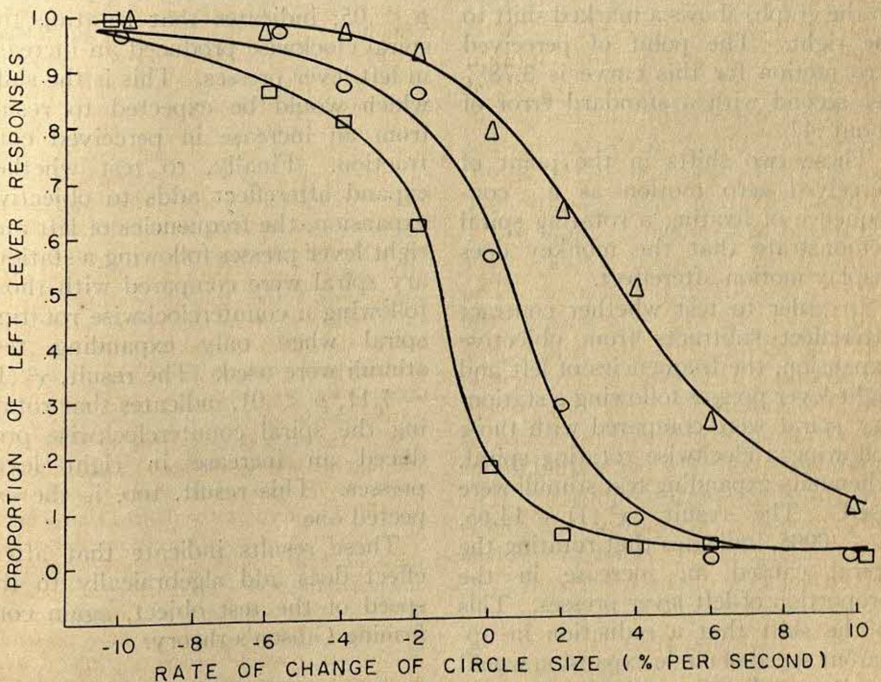


FIG. 1. Proportion of left lever responses as a function of rate of change of circle size for three different conditions of pre-exposure to spirals: stationary spiral (circles on graph), counterclockwise rotation (squares), and clockwise rotation (triangles).

ing was .21% per second with an estimated standard error of .32. Thus, the point of perceived zero motion differs less than one standard error from the point representing objective zero motion.

Inspection of the results from the counterclockwise condition (squares on the graph) shows a systematic shift to the left. For this curve the point of perceived zero motion is at −1.92% per second, with a standard error of about .32. This means that a circle objectively contracting at 2% per second was responded to as if it were motionless. The spiral used was drawn so that counterclockwise rotation would be expected to produce an aftereffect of expansion. It evidently did produce such an aftereffect whose average rate was 1.92% per second.

Inspection of the results from the clockwise rotation condition (triangles on the graph) shows a marked shift to the right. The point of perceived zero motion for this curve is 3.78% per second with a standard error of about .42.

These two shifts in the point of perceived zero motion as a consequence of fixating a rotating spiral demonstrate that the monkey does display motion aftereffect.

In order to test whether contract aftereffect subtracts from objective expansion, the frequencies of left and right lever presses following a stationary spiral were compared with those following a clockwise rotating spiral, when only expanding test stimuli were used. The result, χ^2 (1) = 14.65, $p < .0005$, indicates that rotating the spiral caused an increase in the proportion of left lever presses. This is the shift that a reduction in apparent speed of circle expansion would produce. To test whether expand aftereffect subtracts from objective contraction, the frequencies of left and right lever presses following a stationary spiral were compared with those following a counterclockwise rotating spiral, when only contracting test stimuli were used. The result, χ^2 (1) = 9.90, $p < .005$, indicates that rotating the spiral counterclockwise produced an increase in right lever presses. These two comparisons support algebraic additivity when the aftereffect is opposite in direction to the test motion.

Two similar comparisons can be made to test for additivity when aftereffect and test motion are in the *same* direction. To test whether contract aftereffect adds to objective contraction, the frequencies of left and right lever presses following a stationary spiral were compared with those following a clockwise rotating spiral, when only contracting test stimuli were used. The result, χ^2 (1) = 4.32, $p < .05$, indicates that rotating the spiral clockwise produced an increase in left lever presses. This is the shift which would be expected to result from an increase in perceived contraction. Finally, to test whether expand aftereffect adds to objective expansion, the frequencies of left and right lever presses following a stationary spiral were compared with those following a counterclockwise rotating spiral when only expanding test stimuli were used. The result, χ^2 (1) = 7.11, $p < .01$, indicates that rotating the spiral counterclockwise produced an increase in right lever presses. This result, too, is the expected one.

These results indicate that aftereffect does add algebraically to the speed of the test object, again confirming Gibson's theory.

DISCUSSION

From normative work done in our laboratory on the various parameters of visual aftereffect of motion, it has been found that maximum aftereffect is generally associated with stimulating speeds of about 114′ to 228′ of retinal arc per second. Actual measurement of the speed of the aftereffect places it at about 5′ to 6′ of arc per second (Scott et al., 1963; Scott & Medlin, 1962). The stimulating speeds used by Carlson are in the general range which could be expected to elicit maximum aftereffect. However, the test speeds he used were of the order of 15 times as fast as the aftereffect to be measured. It would be quite surprising if aftereffect would have a measurable influence on the apparent speed of stimuli moving this fast.

Since the judgments made by Carlson's *S*s involved a combination of time and speed estimation, it is possible that the positive effects he did obtain were not due to aftereffect of motion, but resulted from a distortion of subjective time estimation as a consequence of excitatory visual stimulation.

The results of this study appear to leave Gibson's theory of adaptation unchallenged.

REFERENCES

CARLSON, V. R. Adaptation in the perception of visual velocity. *J. exp. Psychol.*, 1962, **64**, 192–197.

GIBSON, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 456–501.

SCOTT, T. R., BRAGG, R. A., & SMARR, R. G. Brain damage diagnosis with the MMG. *J. consult. Psychol.*, 1963, **27**, 45–53.

SCOTT, T. R., & MEDLIN, R. E. Psychophysical measurement of the spiral aftereffect: The MMG. *Amer. J. Psychol.*, 1962, **75**, 319–321.

SPEARMAN, C. The method of "right and wrong cases" without Gauss' formulae. *Brit. J. Psychol.*, 1908, **2**, 227–242.

(Received December 10, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 506–513

EFFECTS OF EXPOSURE VARIABLES ON FIGURAL AFTEREFFECTS UNDER TACHISTOSCOPIC PRESENTATION[1]

ELIZABETH FEHRER AND DONALD GANCHROW

Brooklyn College of the City University of New York

Figural aftereffects with 3 of the Köhler and Wallach stimulus pairs were measured under tachistoscopic exposures. Inspection figure durations were varied from 5 to 1,700 msec. and interstimulus intervals from 0 to 1,000 msec. Test figure exposures were held constant at 15 msec. With the shorter inspection periods, the curves relating magnitude of figural aftereffect to interstimulus interval were in the shape of an inverted U. With 2 stimulus pairs, the longest inspection periods resulted in monotonic curves, the magnitude of the effect decreasing with increase in interval. With the 3rd stimulus pair, all curves were U shaped. In general, the data indicate a reciprocal relation between inspection period and interstimulus interval for the production of maximal aftereffects. This relation is the same as one of Korte's laws for optimal phi, and suggests that brief exposure figural aftereffects and phi may have a similar physiological basis. The Köhler and Wallach and the Osgood and Heyer theories do not seem able to explain the U shaped curves and certain other anomalous findings.

A number of previous studies of figural aftereffects (FAE) have been specifically concerned with the effects of exposure variables. It has been shown that the magnitude of FAE increases with exposure duration of the inspection (I) figure, (e.g., Hammer, 1949; Parducci & Brookshire, 1956; Sagara & Oyama, 1957), and decreases with exposure duration of the test (T) figure (e.g., Krauskopf, 1954; Parducci & Brookshire, 1956). The effect of interstimulus interval (ISI) is less clear-cut. Hammer has shown that FAE decreases as ISI is increased from 5 to 150 sec. Ikeda and Obonai (1953) found similar results for the range of 0 to 90 sec. Parducci and Brookshire, on the other hand, found no differences between 3-sec. and 10-sec. intervals. For very brief intervals, 0 to 500 msec., Ikeda and Obonai (1955) have shown that the relation between ISI and magnitude of FAE is not monotonic. Instead, FAE increased to a maximum at an ISI of 60 msec. and then decreased as the interval was prolonged. In this experiment, the exposure duration of both I and T figures was 500 msec.

The experiments to be reported show (as the experiment of Ikeda and Obonai did for ISI) that when exposure durations and ISIs are very brief, the functions relating FAE to exposure variables are not necessarily monotonic and that therefore the trends that have been observed with longer exposures may show reversals as exposure durations are decreased.

EXPERIMENT I

Method

Apparatus.—The exposure device was a Karlin three-channel tachistoscope (Karlin, 1955) that permitted independent control of the duration of the I and T figures and of ISI.

Two pairs of I and T figures were investigated, Köhler and Wallach's (1944) Fig. 55 and a modification of their Fig. 57.

---

[1] This research was supported by Research Grant G-19199 from the National Science Foundation.

Each figure was drawn in India ink outline, $\frac{1}{8}$ in. thick, on a white card. The dimensions are given in Fig. 1. The field which served as forefield, afterfield, and interstimulus field contained a white card with a fixation dot in the center. The luminance of the white cards was approximately 1 ft-L. The distance from *S*'s eyes to the stimulus cards was 32 in. The maximum dimension of the trapezoid pair subtended 4°40′ and the bent line pair 4°30′.

The magnitude of FAE was measured by having *S* designate that one of a series of matching figures which matched the apparent shape of the T figure. These series were constructed as follows.

For the trapezoid pair, the matching figures consisted of eight pairs of lines. In successive pairs, the separation of the tops of the lines increased by $\frac{1}{8}$ in. (14′ of arc) while the separation of the bottoms of the lines decreased by the same amount. The separation at the centers, 14/16 in., remained constant.

For the bent line pair, the matching figures were 11 bent lines. The two top points of each line were successively raised by $\frac{1}{16}$ in. (7′).

The sets of matching figures were numbered consecutively, the number 1 representing a matching figure exactly like the test figure, and successive numbers representing increasingly deviant figures. Actually, an equal number of matching figures in the opposite orientation appeared on the matching sheets, but these were never selected by *S*s.

Procedure.—The experimental variables were duration of I and of ISI. The exposure of the test figure was kept constant at 15 msec. since with longer exposure the shape of the figure appeared visibly to shift towards the objective shape (cf., Krauskopf, 1954). Since it was not possible to match an apparently changing test figure with a particular matching figure, this constant brief exposure was necessary.

Four I durations, 5, 15, 100, and 500 msec., were explored at each of the six following ISIs: 0, 35, 75, 150, 300, and 1,000 msec. Twenty observations were made at each of these 24 I-ISI duration combinations according to the following procedure. Shuffled orders of the 24 combinations were prepared. Four such sequences of 24 observations each were run on each experimental day, two with each of the two pairs of figures. Forty sequences thus yielded 20 observations for each I-ISI combination for each of the pairs of figures.

The stimuli were presented at 15-sec. intervals following a verbal "ready, one, two, three" signal. At the word "ready," *S* fixated the fixation point. After the exposure, *S* looked at the matching series affixed to the adjoining wall 19 in. to the right of the eyepiece of the tachistoscope, announced his choice, and then looked down at the table top until the next ready signal. The room lights were on throughout.

A preliminary investigation indicated that there was no build-up of FAE with a 15-sec. interval between observations. In this preliminary study in which the trapezoid pair and Köhler and Wallach's Fig. 41 were used, 48 series of 10 observations each were run at four different I durations: 5, 15, 100, and 1,700 msec., and three ISIs: 0, 75, and 1,600 msec. Three naive *S*s took part. In half of these series, the orientation of the I figure remained the same for the 10 observations, whereas in the other half of the series, the orientation of the I figure was alternated

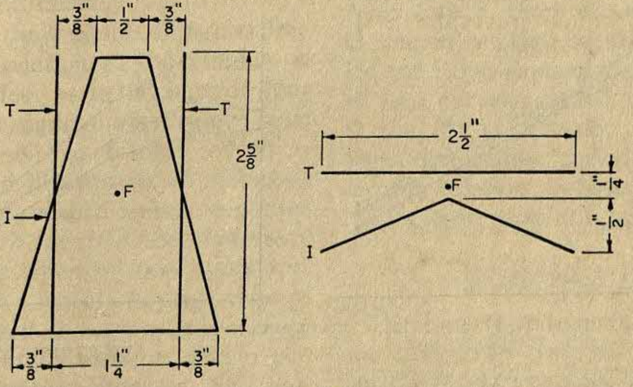


FIG. 1. Trapezoid and bent line pairs. (I represents the inspection figure; T the test figure; F the fixation point.)

in successive exposures. Since the latter procedure should result in successive reversals of the direction of the FAE, one would anticipate less FAE in the alternating series if build-up were a factor. Magnitude of FAE, however, was found to be virtually the same in the two types of series. Since this preliminary study yielded no evidence of build-up, only one orientation of each inspection figure was used in the main experiments.

Subjects.—The authors, one graduate student experienced in tachistoscopic observation, and three naive undergraduates served as *S*s.

Results

The magnitude of FAE was measured by the number of the matching figure chosen. Although *S*s varied in the magnitude of effect, the relations between amount of effect and I and ISI durations were quite comparable. The means for the six *S*s are presented in the upper curves of Fig. 2. Each point is the mean of 120 observations (6 *S*s × 20 trials). The ordinates represent the amount of displacement in units of 1/16 in. (7′ of arc). For the trapezoid pair, the ordinate values indicate the amount of deviation of the upper tip of each of the lines from the parallel position. For the bent line pair, the values indicate the elevation of the right and left upper tips of the apparently bent test line above the center point.

Trapezoid-parallel lines.—The upper left hand curves of Fig. 2 show that the effect of ISI on FAE varied with the duration of I. The magnitude of FAE decreased with increase in ISI for the two longest inspection periods. This trend is continuous with that reported in previous studies in which the ISIs ranged from 1 sec. up. For the 5-msec. I exposure, however, the relation between magnitude of FAE and ISI was inverted **U** shaped, the

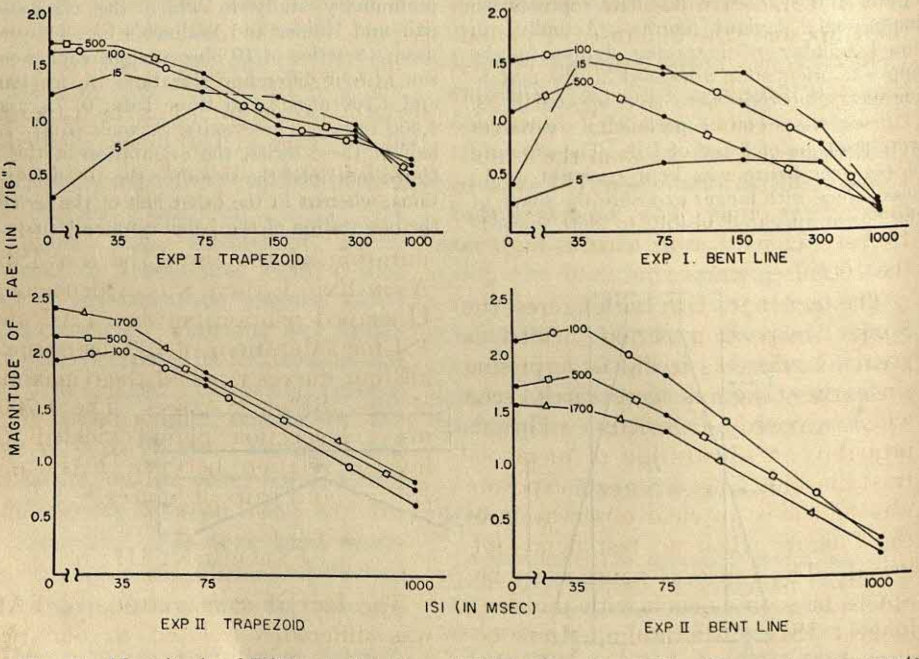


FIG. 2. Magnitude of FAE as a function of ISI in Exp. I and II. (Parameter represents duration of inspection figure in milliseconds; ISI values are plotted according to the transformation $x' = \log_2 (x/18.75)$, when $x \geq 35$.)

effect maximizing at an ISI of 75 msec. For the 15-msec. I duration, there was also a suggestion of a U shaped relation.

The effect of I duration on magnitude of FAE, therefore, depended on ISI. When ISI was 0 or 35 msec., FAE increased with I duration. With ISIs of 300 msec. and beyond, FAE was independent of I duration. For this pair of figures, an increase in I duration from 100 to 500 msec. had very little effect on FAE, the two curves being virtually identical.

Bent line-straight line.—The U shaped relation between ISI and FAE was more pronounced for this pair of figures. The 5- and 15-msec. I duration curves are clearly U shaped and the 100- and 500-msec. curves very slightly so.

As with the trapezoid pair, the duration of the inspection period affected FAE differentially only at the briefer ISIs. Unlike the results for the trapezoid, the 500-msec. I period yielded consistently smaller effects than the 100-msec. I period. Both of these inspection periods, however, yielded greater effects than the 5- and 15-msec. I periods when ISI was 0 msec. At longer ISIs, however, the 15-msec. curve was at least as high as the 100-msec. curve.

The fact that, with both figures, the 5-msec. inspection period yielded as great an effect as it did is surprising since the I figures were rarely seen when exposed so briefly. This is probably an illustration of metacontrast masking since a 5-msec. exposure was sufficient for clear observation of the I figure when no test figure followed. The 5-msec. I figure was more apt to be seen as such with the three longest ISIs. Maximum effects occurred, therefore, at those ISIs that marked the transition between "seeing" and "not seeing" the I figure.

EXPERIMENT II

The finding that a 500-msec. inspection period yielded little if any more FAE than the 100-msec. duration with the trapezoid pair and a distinctly smaller effect with the bent line pair, led us to investigate the effect of an even longer inspection period.

Method

The authors and the graduate student served as *S*s. The I durations investigated were 100, 500, and 1,700 msec., the last being the longest pulsed duration that the tachistoscope permitted. The ISIs studied were 0, 35, 75, and 100 msec. Twenty observations were recorded for each I-ISI combination as in Exp. I. The data appear in the lower half of Fig. 2. The fact that the magnitudes of FAE are greater than in Exp. I is due to *S* differences. The three *S*s used in Exp. II had shown greater effects than the naive *S*s of Exp. I. Actually, the FAEs for the inspection periods common to both experiments were very similar for these three *S*s.

Results

The results for the two figures are distinctly different. For the trapezoid pair, increase in I duration yielded a slight increase in FAE at the briefer ISIs. For the bent line pair, the effect was inversely related to I duration at the three shortest ISIs. As in Exp. I, there was evidence of a U shaped relation between FAE and ISI for I durations of 100 msec. since all four curves reached their maxima at the ISI of 35 msec. The 1,700-msec. inspection period yielded an inverse relation between FAE and ISI for both pairs of figures.

EXPERIMENT III

The fact that magnitude of FAE was differently related to our two experimental variables in the two pairs of figures selected for investigation led us to study one additional

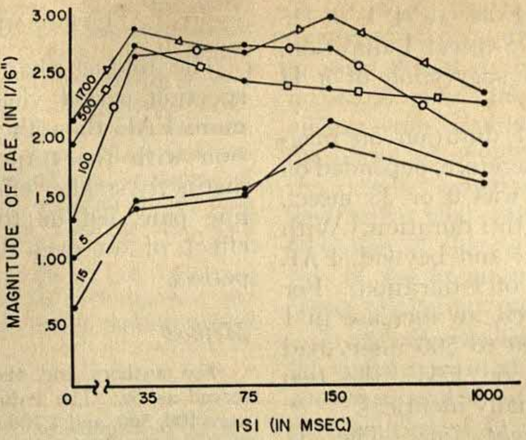


FIG. 3. Magnitude of FAE as a function of ISI in Exp. III.

pair of figures, namely, concentric circles. This pair was chosen so that we could compare our results with those of Ikeda and Obonai.

Method

The authors and the graduate student served as *S*s. The I figure was a 2-in. diameter circle with the center 1¾ in. (3°10′) to the right of the fixation point. The test and comparison circles were 1¼ in. in diameter, the former concentric with the I circle and the latter at an equal distance to the left of the fixation point. In the matching series, the comparison circle in each pair remained constant in size while the radius of each successive test circle decreased by $\frac{1}{16}$ in. After each observation, *S* designated the number of the pair whose size ratio matched his observation in the tachistoscope.

The I durations used were 5, 15, 100, 500, and 1,700 msec.; the ISI durations, 0, 35, 75, 150, and 1,000 msec. The procedure was the same as in Exp. I.

Results

The mean values, appearing in Fig. 3, show a **U** shaped relation between magnitude of FAE and ISI for all durations of I. This finding differs from that for the trapezoid and bent line pairs, since for these the relation between ISI and FAE was monotonic following the long inspection periods. Our findings are in rough agreement with those of Ikeda and Obonai (1955) who also found a **U** shaped relation between FAE and ISI after a 500-msec. inspection period.

For the circles, the relations between I duration and FAE were also different in certain other respects from those for the trapezoid and bent line pairs. For the concentric circles, the 15-msec. exposure was, if anything, less effective than the 5-msec. exposure. As with the trapezoid, the longest inspection period yielded the greatest effects. Finally, I duration was an effective variable at the ISI of 1,000 msec. whereas for the other two figures FAE was independent of I duration at this ISI.

DISCUSSION

The data described above show clearly that the effects of duration parameters on magnitude of FAE are not always monotonic over a range of extremely brief durations. The relations are not, therefore, continuous with those that have been found for durations of the order of seconds and minutes as in the study by Hammer (1949). Theories such as those of Köhler and Wallach (1944) and Osgood and Heyer (1952), which require an increase of FAE with increase in I duration, and a decrease of FAE with increase of ISI, cannot explain the numerous exceptions to these generalizations in our data. Further, the magnitude of some of our FAEs, up to 20′ of arc, is probably greater than the Osgood and Heyer theory can account for (cf., Krauskopf, 1954).

It seems far more probable that the physiological basis for our brief exposure FAEs is related to that underlying apparent movement. First, both effects, (FAE and phi), occur under the same experimental conditions. Our stimulus durations and intervals included the range typically investigated in studies of phi, and our procedure was similar to the single sequence procedure used in many investigations of phi (cf., Neff's 1936 review). Whether phi is typically accompanied by displacement of the second stimulus is not known since studies of phi have not been designed to measure such displacement. However, individual protocols in some of these studies (e.g., Higginson, 1926, pp. 87–88), indicate that such displacement does occur.

Second, our data indicate that the effect of our duration variables on FAE is similar to the effect of the same variables on phi, and, in fact, follows that one of Korte's laws which states an inverse relation between exposure duration and ISI for optimal phi (cf., Graham, 1951). Thus, maximum FAEs required longer ISIs with short inspection periods (5 and 15 msec.) than with longer ones (100 msec. and beyond), and, similarly, optimum phi requires longer ISIs when presentation times are brief. Furthermore, our finding with the trapezoid and the bent line pairs that optimum FAE occurred at an ISI of 0 msec. when the inspection period was long is consistent with the fact that good phi results when a second brief stimulus is presented simultaneously with the end or slightly overlapping the end of a long duration first stimulus (cf., Higginson, 1926; McConnell, 1927).

Additional evidence linking brief exposure FAEs with apparent movement comes from a statement of Ikeda and Obonai (1955) on the relation between magnitude of FAE and the impression of successiveness. They pointed out that the impression of successiveness or of seen movement for their concentric circles first occurred at an ISI of 60 msec., and that this ISI yielded maximum FAE under their exposure conditions. Similarly, we observed that strong FAEs required an impression of successiveness of the two figures but we unfortunately used very few durations that yielded the impression of simultaneity. The following brief experiment, however, showed that maximum FAE does not necessarily occur under that particular combination of conditions that first yields the perception of succession.

Two figure pairs, the trapezoid-parallel lines and the concentric circles, were used. The inspection figures were presented for the following five durations: 15, 50, 100, 500, and 1,700 msec., and followed, after one of the five following ISIs: 0, 20, 40, 60, and 80 msec., by the 15-msec. test figure. A 5-msec. inspection period could not be included since this was too brief to permit "seeing" of the I figure.

The three experienced *S*s judged each of these 25 I-ISI combinations three times each in random order. If the I and T figures seemed simultaneous, a rating of 1 was assigned; a rating of 2 if they seemed to overlap in time (i.e., the I figure was still present when the T figure appeared but disappeared sooner); and a rating of 3 if they appeared clearly successive.

The three judgments of each *S* were highly consistent and the results for the two figures were very alike. The means, based on 18 observations each, (2 figures × 3 *S*s × 3 trials), are presented in Table 1.

It will be observed that an ISI of 80 msec. typically yielded an impression of succession at all I durations. At briefer ISIs, however, the impression of succession was related to I duration. Thus, for ISI = 0 msec., an I figure exposed for 15 msec. seemed simultaneous with the

TABLE 1

EFFECTS OF EXPOSURE VARIABLES ON RATINGS OF SIMULTANEITY OR SUCCESSIVENESS OF I AND T FIGURES

Duration of ISI (in msec.)	Duration of I (in msec.)				
	15	50	100	500	1,700
0	1.0	1.2	1.8	2.6	2.9
20	1.1	1.9	2.4	2.8	3.0
40	1.6	2.4	2.8	3.0	2.9
60	2.6	2.8	2.9	2.9	3.0
80	2.8	2.8	3.0	3.0	3.0

test figure, whereas, if the same I figure was exposed for 1,700 msec., it was seen to disappear before the test figure appeared.

The fact that the appearance of succession required a longer ISI with short than with long inspection periods at first seems to be a surprising fact. One would anticipate that a longer lasting stimulus would yield a more persistent aftereffect and that, therefore, a longer ISI would be required for the appearance of succession. These findings are, however, quite consistent with Korte's law since, over a limited range of durations, there was reciprocity between exposure time and ISI for the first appearance of successiveness.

A comparison of the data in Table 1 with those in Fig. 2 and 3 shows a number of exceptions to Ikeda and Obonai's statement. For example, for the trapezoid and bent line pairs, a 15-msec. inspection period yielded maximum FAE when ISI was 35 msec., whereas successiveness was not clearly established at this interval. For the bent line figures, FAE also maximized at an ISI of 35 msec. when the inspection period was 500 msec., although successiveness was already present when ISI was 0 msec. Finally, although the trapezoid and concentric circle pairs of figures yielded almost identical impressions of simultaneity or succession at the same exposure combinations, the curves relating FAE to exposure variables were quite different. Although our data do not show a precise relation between maximum FAE and the appearance of succession, they do not contradict the statement that the experience of successiveness is necessary for large FAEs. Certainly, the 15-msec. exposure of all inspection figures yielded less than maximal FAE when ISI was 0 msec., and at this ISI inspection and test figures always appeared to be simultaneous.

If the visual FAEs found under brief exposure conditions should prove to be based on the same physiological mechanisms as phi, the question remains whether FAEs found under long exposure conditions as in the Köhler and Wallach experiments can be similarly explained. It is possible that the mechanism underlying the latter FAEs is different. If so, however, long exposure FAEs are qualitatively similar to those found with brief exposures. This was clearly true of the three sets of figures reported here and also of the FAEs with six other of the Köhler and Wallach figures which we have observed under our exposure conditions. Aftereffects that persist for days in the absence of stimulation with the I figure (cf., McEwen's 1958 review, p. 78) certainly seem to require principles in addition to those required for the explanation of phi.

It should, perhaps, be pointed out that tactual and visual FAEs do not seem to be similarly related to ISI. Helson and King (1931), in their study of the tau effect, found that the apparent tactual distance between two points increased as ISI increased from 200 to 500 msec. We, on the other hand, found that the apparent distance between the two relevant lines of the I and T figures decreased (less FAE) as ISI increased beyond 150 msec. in the case of brief inspection stimuli and beyond 75 msec. for longer ones. This discrepancy may, however, be more apparent than real since the nature of the stimuli (points vs. lines) was different, and since the ISIs for optimum FAEs may be quite different for skin and retina.

We can offer no good explanation for the differences found between figures.

It is not clear why, with the concentric circles, all curves relating ISI to FAE were U shaped whereas with the other two figures only the briefer inspection periods yielded U shaped curves. Neither can we explain the fact that, with the bent line pair, an increase in the inspection period beyond 100 msec. resulted in a decrease in FAE. The ultimate explanation may involve configurational principles; it may depend on details of the positioning of the I and T lines; or it may be related to the amount of overlap of the I and T figures.

REFERENCES

GRAHAM, C. H. Visual perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

HAMMER, E. R. Temporal factors in figural after-effects. *Amer. J. Psychol.*, 1949, 62, 337–354.

HELSON, H., & KING, S. M. The tau effect: An example of psychological relativity. *J. exp. Psychol.*, 1931, **14**, 202–217.

HIGGINSON, G. D. The visual apprehension of movement under successive retinal excitations. *Amer. J. Psychol.*, 1926, **37**, 63–115.

IKEDA, H., & OBONAI, T. The quantitative analysis of figural after-effects: I. The process of growth and decay of figural after-effects. *Jap. J. Psychol.*, 1953, **23**, 246–260.

IKEDA, H., & OBONAI, T. The studies in figural after-effects: IV. The contrast-confluence illusion of concentric circles and the figural after-effect. *Jap. psychol. Res.*, 1955, **2**, 17–23.

KARLIN, L. The New York University tachistoscope. *Amer. J. Psychol.*, 1955, **68**, 462–466.

KÖHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

KRAUSKOPF, J. The magnitude of figural after-effects as a function of the duration of the test period. *Amer. J. Psychol.*, 1954, **67**, 684–690.

MCCONNELL, R. F. Visual movement under simultaneous excitations with initial and terminal overlap. *J. exp. Psychol.*, 1927, **10**, 227–246.

MCEWEN, P. Figural after-effects. *Brit. J. Psychol. monogr. Suppl.*, 1958, No. 31.

NEFF, W. S. A critical investigation of the visual apprehension of movement. *Amer. J. Psychol.*, 1936, **48**, 1–42.

OSGOOD, C. E., & HEYER, A. W., JR. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.

PARDUCCI, A., & BROOKSHIRE, K. Figural after-effects with tachistoscopic presentation. *Amer. J. Psychol.*, 1956, **69**, 635–639.

SAGARA, M., & OYAMA, T. Experimental studies on figural after-effects in Japan. *Psychol. Bull.*, 1957, **54**, 327–338.

(Received December 13, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 514–520

INFLUENCE OF TYPE OF DISCRIMINATION TRAINING ON GENERALIZATION[1]

KENNETH L. HOVING[2]

University of Washington

An attempt is made to evaluate discrimination procedures which differed either as to the availability of reinforcement or in terms of the relationship of the stimulus dimension used in training to that employed in tests of generalization. Initially 20 college students were used to determine an equal loudness contour for the 4 tones later used in generalization testing. Tests of generalization across pitch were conducted immediately following discrimination training. Nonparametric analysis of the operant data of generalization testing for 3 groups of 16 Ss each reveals: (a) that the relationship of the stimulus dimension used in training to that employed in testing is an important determiner of generalization, (b) that negation of control by incidental stimuli does not always result in sharply sloping gradients of generalization, (c) that the response rate emitted in testing is composed of varying portions of the rates emitted in training.

One of the parameters of stimulus generalization identified by both classical conditioning studies and the operant approach has been that of discrimination training. Studies by Reingold and Perkins (1955), Perkins, Hershberger, and Weyant (1959), Butter (1959), and Jenkins and Harrison (1960) suggest that discrimination training decreases stimulus generalization. Hull (1943, 1950, 1952) has suggested two mechanisms to account for these findings. One, also suggested by Spence (1936, 1942), involves the algebraic summation of excitatory and inhibitory gradients generated by the reinforced and nonreinforced stimuli, respectively. A second mechanism suggested by Hull (1950, 1952) involves the role of stimuli arising from certain constant features of the apparatus. When training is nondifferential, these stimuli are always present when reinforcement occurs and may predominate in control of the response with the result that the gradient of generalization observed upon varying the experimental stimuli is flat or nearly so. The present study is an attempt to shed further light on the operation and applicability of these mechanisms in human *S*s by comparing the effect of discrimination training involving reinforcement and nonreinforcement with training in which reinforcement was always available contingent upon the appropriate response.

A second goal was the investigation of the effect of discrimination training on stimulus generalization when training and testing involve differing dimensions of the same stimulus. Stevens and Davis (1938) have suggested that the loudness and pitch dimensions of tone can be varied independently. The loudness of tone was selected as the dimension to be used in discrimination training with tests of stimulus generalization being conducted across pitch with loudness remaining constant. Unfortunately, these psychological dimensions are not linearly related to the physical dimensions of intensity and frequency. Available equal loudness contours were derived using techniques which differed markedly from those necessary in the present study. Therefore, the first step in the present study was the derivation of an equal loudness level across those frequencies which would be used to test for stimulus generalization.

---

[1] This report is based on part of a PhD dissertation presented to the Graduate School of the University of Washington. The author is deeply indebted to Moncrieff H. Smith, Jr. for his advice and encouragement.

[2] Now a National Institute of Mental Health Postdoctoral Research Fellow at the University of Minnesota.

PRELIMINARY EXPERIMENT

Method

Subjects.—Twenty *S*s between 18 and 30 yr. of age from an upper division psychology course were used in the derivation of the equal loudness curve. All *S*s reported normal hearing.

Apparatus.—Auditory stimuli were produced by a fixed frequency oscillator (Heathkit, Model AG-9A) and were delivered via dynamic earphones (Permaflux, Model PDR-10). The oscillator output was monitored via a voltmeter.

Procedure.—The *S* was seated in a large easy chair and was instructed "to judge the loudness—not pitch but loudness—of tones." The *S* was then handed the earphones and informed that he would hear tones only through the left ear. The *E* attempted to insure uniformity of earphone placement by *S*. A standard tone, 1,000 cps at 90 db., was presented prior to each comparison tone. Each tone was presented for 30 sec. with a 15-sec. interval between tones. The order in which the comparison tones were presented was determined by a table of random numbers. The intensity at which the tone was first presented was randomly varied between three intensity levels, 86, 90, and 94 db. If the initial judgment was other than same, the intensity of the comparison tone was changed in the direction opposite the judgment. This process was continued until a judgment of same was obtained for that tone. Each *S* was given the same randomized series of test tones twice.

Results and Discussion

The results are shown in Table 1. The data were also subjected to an analysis of variance for repeated measures. This analysis indicated that tones in this limited frequency range must differ significantly in intensity, $F(3, 57) = 4.11$, $p < .05$, to be judged as loud as the standard (1,000-cps tone at 90 db.). That an equal loudness curve over this span should necessitate significant differences in intensity is somewhat surprising, especially when these results are compared with those reported in the literature. It must be kept in mind, however, that the technique by which the present curves were derived differed markedly from those used by Fletcher and Munson (1933) and Stevens and Davis (1938). In addition, the present technique forces *S*s to make comparisons under conditions which, although identical to those used in the later experiment, are nevertheless less than optimal for generating accurate absolute equal loudness curves. The significant difference among *S*s, $F(19, 57) = 2.78$, $p < .05$, is also not too surprising as the variance contributed by the order in which the comparison tones were presented is included in the *S* variance.

TABLE 1

INTENSITY AT WHICH TONES WERE JUDGED TO BE AS LOUD AS A 1,000-cps TONE AT 90 db.

Frequency (cps)	Intensity (db.)	
	M	*SD*
590	89.95	2.00
770	90.67	2.22
1,000	90.00	0.00
1,300	91.00	2.56
1,700	92.30	3.36

Note.—Average of two judgments by each of 20 *S*s.

MAIN EXPERIMENT

Method

Subjects.—Forty-eight students enrolled in lower division psychology courses and between the ages of 18 and 30 were used as *S*s. The *S*s were randomly divided into one of three experimental groups.

Apparatus.—The apparatus was identical to that used in the determination of an equal loudness curve with the following additions: A telegraph key was attached to the right arm of the chair in which *S* was seated. A small neon light located 3 in. in front of the manipulandum served to provide reinforcers. A red light was placed 5 in. in front of the reinforcement light. Responses were recorded on a Gerbrands cumulative recorder in which 14.5 responses displaced the pen laterally 1 cm. The magnitude of this excursion enabled *E* to observe readily the form of the response.

Procedure.—Those *S*s who reported normal hearing were read one of three sets of instructions. These instructions were as uniform as possible in light of the differences in discrimination training. In essence, the instructions informed *S* he would have to learn at what rates (Groups I and III) or rate (Group II) to press the bar for tones (Groups I and II) or tone and light (Group III) to make the reinforcement light go on.

Discrimination training.—Groups I and II were given discrimination training on the loudness dimension of a 1,000-cps tone. Responses emitted in the presence of Stimulus S_1, a 1,000-cps tone at 90 db., were reinforced by a gradually increasing variable ratio. Manual shaping of the response rate was initially used to achieve a moderately fast rate. Responses to Stimulus S_2, a 1,000-cps tone at 80 db., were reinforced in Group I on a DRL variable ratio schedule. Reinforcements were only delivered if responses were separated by a 2-sec. interval. Group II was never reinforced for responding to Stimulus S_2. In Group III responses emitted in the presence of a 1,000-cps tone presented at both 80 and 90 db. were reinforced on the same VR schedule as that used for Stimulus S_1 in Groups I and II. These two intensity levels were presented equally often in training and were treated as Stimulus S_1. Stimulus S_2 was a 7-w. red light. No tone was presented when the red light, Stimulus S_2, was on. Responses emitted in the presence of this stimulus were reinforced on the same DRL VR schedule employed for Stimulus S_2 in Group I.

The discrimination procedure initially involved the alternate presentation of Stimuli S_1 and S_2 until stimulus control over response rate was achieved. No control over stimulus duration was exercised during this initial period of training but the interval between presentations remained 15 sec. If stimulus control was not achieved after 16 trials, or more than 20 min. had elapsed, *S* was dropped from the experiment. Once stimulus control was achieved the stimuli were presented for 30 sec. with an interval of 15 sec. between trials. The order of presentation from this point on was governed by use of a table of randomized trial series developed by Gellerman (1933). Stimuli S_1 and S_2 were randomly presented with the restriction that the same stimulus not occur more than three times in succession and not more than five times in 10 trials.

This procedure was continued until the following three discrimination criterion measures were met for Group I: that the ratio of responses was at least 7:1 during the presentation of Stimuli S_1 and S_2, respectively; that this ratio be maintained for at least three cycles of S_1 and S_2; and that no change in rate of responding occur when reinforcement was deleted during the presentation of both S_1 and S_2 on two trials.

The discrimination criterion for Group II was that no response be emitted in the presence of Stimulus S_2 for three consecutive cycles of S_1 and S_2; that at least 70 responses be emitted during the presence of Stimulus S_1 on two occasions in which no reinforcement was given.

The discrimination criterion measures were the same for Group I and Group III with one exception, that there be a difference of less than five responses between the number of responses emitted in the presence of Stimulus S_1 at 80 db. and Stimulus S_1 at 90 db. in Group III. If the criterion used in any of the discrimination training procedures was not met within the first 40 min. of the experiment, *S* was dropped. This resulted in the loss of four *S*s in Groups I and II and seven *S*s in Group III. Generalization testing was begun immediately after the criterion was met.

The procedure used to test for generalization was the same for each group. The discriminative stimuli, S_1 and S_2, were both presented 10 times during generalization testing. The order in which these stimuli were presented was the same for each *S*. The four test stimuli were interspersed among the discriminative stimuli. The order in which the test stimuli were presented was the same as that used for the first 16 *S*s in deriving the equal loudness curve. These 16 randomly determined orders were used in each of the three experimental groups. Each test cycle was presented twice to each *S* with each test stimulus being presented immediately after both discriminative stimuli, S_1 and S_2.

This procedure was used to equalize any effects due to the order of presentation. The test stimuli differed from the training stimuli in frequency but were presented at an intensity judged equal to that of a 1,000-cps tone at 90 db. Responses emitted in the presence of the test tones were never reinforced. An equal number of reinforcements was delivered in the presence of Stimuli S_1 and S_2 during testing.

Results

The generalization gradients obtained following these three types of discrimination training are shown in Fig. 1. Each test point on the gradient represents the average number of responses emitted in two 30-sec. intervals by 16 *S*s. Because of the variability in the generalization gradients shown by individual *S*s, one could not assume homogeneity of variance or normalize the data with a transformation; hence, nonparametric statistics were used in the analysis of the results. Friedman's two-way analysis of variance for related samples (Siegel, 1956) indicates that there are significant differences in the rate of response emitted in each experimental group across their respective test stimuli and to Training Stimulus S_1. In fact, this test allows us to reject the null hypothesis of equal rate at each point on the continuum for each group at the .001 level of confidence.

Table 2 contains the average number of responses emitted by the 16 *S*s in each group at each point on the continuum. A Kruskal-Wallis one-way analysis of variance by rank (Siegel, 1956) was performed for each stimulus on the continuum with the exception of Stimulus S_2. This analysis revealed that the three groups differed significantly at each point

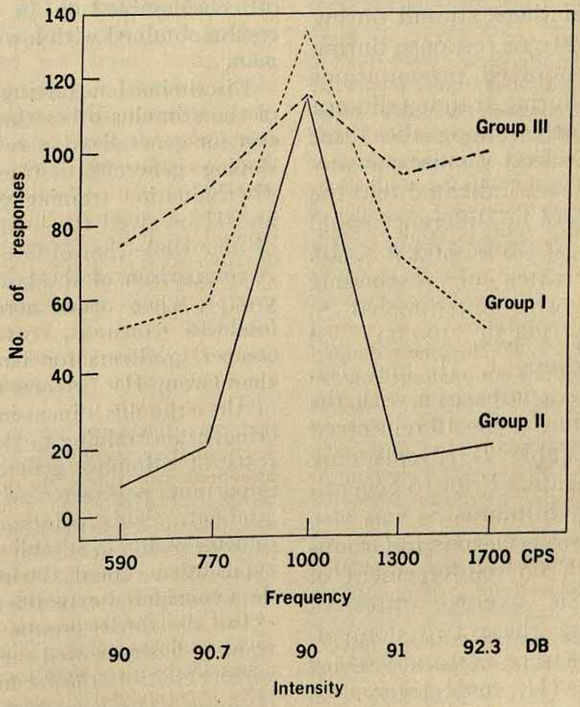


FIG. 1. Generalization gradients representing the average number of responses for 16 *S*s in each group.

TABLE 2

AVERAGE NUMBER OF RESPONSES EMITTED AT EACH POINT ON THE FREQUENCY CONTINUUM

Group	590 cps	770 cps	S_1	1,300 cps	1,700 cps	S_2
I	53.44	58.78	133.96[a]	73.72	57.34	13.37[a]
II	8.97	23.56	116.73[a]	16.41	21.66	3.97
III	75.37	89.94	118.17[a]	96.34	99.09	9.00[a]

[a] Average of 10 reinforced trials by each *S*.

with the one exception being S_1, the training stimulus. These four differences were significant at the .001 level. No comparison of response rate emitted in the presence of Stimulus S_2 was made due to the instruction, criterion, and reinforcement differences used in the training of the three groups.

The average number of responses emitted by each group at Stimuli S_1 and S_2 was derived by averaging the response rate of each *S* to the 10 presentations of these stimuli during testing. The rate of response during the two nonreinforced presentations of Stimulus S_1 during training did not differ significantly among the three groups. A Kruskal-Wallis one-way analysis of variance indicated that the null hypothesis of no difference could not be rejected, $H\ (2) = 4.63, p < .10$. The average rates of responding during these two nonreinforced S_1 intervals were 134.12, 122.90, and 111.46 for Groups I, II, and III, respectively. A comparison with the response rate during the 10 reinforced S_1 intervals (Table 2) reveals only slight shifts in rate. Prior to generalization testing, Stimulus S_2 was also presented on two occasions to Groups I and III with no reinforcement of responses. The average rates of response during these two nonreinforced intervals were 8.50 and 9.44 for Groups I and III, respectively. A comparison of these rates with that obtained during the 10 reinforced trials shown in Table 2 again reveals only slight changes.

Discussion

The results reveal that the three gradients are bidirectional, approximately symmetrical, and with the modal response rate at the training stimulus, S_1. These results, obtained with human *S*s, are in general agreement with similar studies employing nonverbal organisms. The present technique, therefore, may be of considerable use in extending the results obtained with lower organisms to man.

Discrimination training on a dimension of the stimulus other than that used to test for generalization results in sharply sloping generalization gradients. The discrimination training given Groups I and II involved the loudness dimension of tone while that of Group III did not. A comparison of the gradients of these groups, when tested across pitch, with loudness constant, reveals markedly steeper gradients for Groups I and II than Group III. Hence, the relationship of the stimulus dimension used in discrimination training to that employed in tests of stimulus generalization is an important parameter of the resultant gradient. This points out the multidimensionality of stimuli which, although typically accepted theoretically, is seldom considered experimentally.

In all three groups discrimination training has narrowed the class of stimuli which control the occurrence of the response. However, in Groups I and II this training has narrowed the control to one dimension of the auditory stimulus, whereas in Group III no such awareness of, or attending to, the dimensionality of tone is necessary. In fact, by reinforcing responses on the same schedule in the presence of both an 80- and 90-db. tone, S_1, we have made the loudness dimension of the auditory stimulus irrelevant in terms of controlling the response in Group III. This difference in awareness of, or attending to, the dimensions of the auditory stimulus would seem to be a primary factor in accounting for the marked differences in generalization.

An alternative explanation of the marked differences in generalization also suggests itself. If the training stimuli, S_1 and S_2, are reinforced during generalization testing, might not the resultant generalization gradients reflect the formation of a new discrimination based on the reinforcement and nonreinforcement of the training and test stimuli, respectively? There are two factors which suggest that such an interpretation is not in keeping with the present findings. First, all three groups had experienced at least two nonreinforced presentations of both Stimuli S_1 and S_2 in training with response rate remaining constant. Both the reinforcement schedule and the 30-sec. trial length used were designed to make the nonreinforcement of the test stimuli as difficult to discriminate as possible. The raw data also reflect that, although there is greater variability on the first test cycle, the form of the generalization gradients is not markedly different when only the first four test trials are considered.

The gently sloping generalization gradient of Group III indicates that a modification of the precise properties of the stimulus (pitch) even when they are irrelevant in discrimination training does result in a gradient which cannot be considered flat. However, the relatively broad generalization which occurs under these conditions is somewhat surprising as discrimination training is thought to neutralize the control of the response by incidental stimuli. Although the present training does narrow the class of stimuli which controls the response, modification of one of these classes results in a generalization gradient which, although not flat, is very broad.

Although the gradients of Groups I and II differ markedly from that of Group III, they also differ from each other. Here the stimulus dimensions used in training and testing were identical; however, the type of discrimination training used differed. Group I was reinforced for responding to both Stimuli S_1 and S_2, albeit for different rates; whereas, Group II was reinforced for responding to Stimulus S_1 but never for responding to S_2. The basic mechanism suggested by classical S-R theory as responsible for the effect of discrimination training on generalization is differential reinforcement. Since only Group II received training of this type, current S-R theory does not enable us to account for the sharply sloping gradient of Group I. However, Hull himself cautioned against the application of his theoretical system to human behavior without empirical verification. The present study may be viewed in part as an attempt to determine the applicability rather than the validity of a portion of Hull's theoretical system as it related to human behavior. In this sense, the comparison of the gradients of Groups I and II suggests the need for a modification of Hull's differential reinforcement mechanism as applied to human behavior in situations of this type.

One of the findings of considerable interest was that the response to the test stimuli was not a new stable rate intermediate between those emitted during training, but rather was composed of varying portions of these rates. This would initially seem to explain the discrepancy between the gradients of Groups I and II. However, although Group I alternated between responding rapidly and slowly, and Group II alternated between a fast rate, equivalent to that of Group I to S_1, and no responding in the presence of the test stimuli, other factors also seem to be involved. The rate of responding during the presentation of S_2 in Group I averaged only 13.37 responses per 30-sec. interval. If both groups in generalization emit an equivalent combination of the rates developed to Stimuli S_1 and S_2 in training, the two gradients would have been separated by less than 13.37 responses at each point on the continuum. As this was clearly not the case, other factors must also be involved. One such factor would seem to be the orientation or set produced by the reinforcement schedule. Group I was always reinforced for responding in training, whereas Group II was not and thus approached generalization testing with a different orientation (an orientation which resulted in very rapid, although atypical, extinction in Group II). This extinction is atypical in the sense that it appears abruptly and frequently generalizes to all points on the continuum. It also seems plausible that the slight differences in the wording of the initial discrimination instructions may be responsible for this atypical extinction. The present study does not enable us to evaluate the influence of these two factors.

REFERENCES

Butter, C. M. Stimulus generalization and discrimination along the dimension of wavelength and angular orientation. Unpublished doctoral dissertation, Duke University, 1959.

Fletcher, H., & Munson, W. A. Loudness, its definition, measurement, and calculation. *J. Acoust. Soc. Amer.*, 1933, **5**, 82–108.

Gellerman, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42**, 206–208.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Hull, C. L. Simple qualitative discrimination learning. *Psychol. Rev.*, 1950, **57**, 303–313.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Jenkins, H., & Harrison, R. Effect of discrimination training on auditory generalization. *J. exp. Psychol.*, 1960, **59**, 246–253.

Perkins, C. C., Hershberger, W. A., & Weyant, R. G. Difficulty of a discrimination as a determiner of subsequent generalization along another dimension. *J. exp. Psychol.*, 1959, **57**, 181–186.

Reingold, D. B., & Perkins, C. C. Stimulus generalization following different methods of training. *J. exp. Psychol.*, 1955, **49**, 423–427.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Spence, K. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

Spence, K. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, **31**, 257–271.

Stevens, S. S., & Davis, H. *Hearing: Its psychology and physiology.* New York: Wiley, 1938.

(Received December 14, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 521–522

SUPPLEMENTARY REPORTS

DIFFERENTIAL MONETARY GAINS AND LOSSES AND EVENT PROBABILITY IN A TWO-CHOICE SITUATION[1]

JEROME L. MYERS, JANE G. FORT, LEONARD KATZ, AND MARY M. SUYDAM
University of Massachusetts

Ss were run under light combinations of π (probability of the more frequent event) and monetary gain and loss for 400 trials in a noncontingent 2-choice situation. All curves rose above π, monetary incentive yielded greater percent prediction of the more frequent event, and 1¢ and 10¢ gains and losses had similar effects. A Markoff model of choice behavior, which assumes that a stimulus element may be weakly or strongly conditioned to each response, was successful in predicting asymptotic response percentages.

Although monetary gains and losses have been employed as consequences for correct and incorrect predictions of events in several noncontingent two-choice experiments, none of these experiments has involved more than one level of incentive at more than one level of π (probability of the more frequent event). The present experiment was performed in order to provide a more adequate representation of the joint effects of incentive and event probability and, consequently, a more adequate test of a mathematical model which has recently been proposed to deal with these variables.

Method.—Nine groups of 20 *S*s[2] were run, representing all combinations of π's of .6, .7, and .8 and gains and losses of 0¢, 1¢, and 10¢. The *S*s were run in groups of three and four (but separated by partitions) for 400 10-sec. trials. All *S*s were informed of their payoff condition and were told to guess which of two lights would occur on each trial. The *S*s in the 1¢ and 10¢ groups kept track of their gains and losses with pencil and paper provided by *E*; those in the 0¢ groups were run under the usual conditions for experiments not involving monetary payoffs.

Results and discussion.—Table 1 presents the mean proportion of choices of the more frequent event (A_1 responses) for each 100-trial block as well as predictions and *SD*s for the last block of trials. The main features of the data are these: even under no monetary incentive percent choice overshoots for the .7 and .8 groups the asymptote of π predicted by several mathematical theories, A_1 responses clearly increase under monetary incentive, and this effect is only slightly greater for 10¢ than for 1¢. The overshooting of π is consistent with the results of experiments involving more than 300 trials (Edwards, 1961, p. 392). The closeness of results under 1¢ and 10¢ suggests that the function relating percent choice and motivation is negatively accelerated, and levels off extremely quickly.

TABLE 1

OBSERVED PROPORTION OF CHOICES OF THE MORE FREQUENT EVENT, ASYMPTOTIC (Trials 301–400) PREDICTIONS FROM ATKINSON'S WEAK-STRONG (W-S) MODEL, AND *SD* OF PROPORTIONS (SD_p) FOR TRIALS 301–400

π	Gain or Loss	Trials				W-S Predictions	SD_p
		1–100	101–200	201–300	301–400		
.6	0¢	.529	.554	.604	.616	.632	.088
.6	1¢	.544	.585	.641	.649	.682	.156
.6	10¢	.529	.597	.653	.713	.710	.143
.7	0¢	.601	.698	.717	.753	.768	.087
.7	1¢	.620	.783	.825	.870	.831	.114
.7	10¢	.641	.817	.856	.868	.881	.088
.8	0¢	.682	.823	.853	.871	.869	.082
.8	1¢	.728	.890	.926	.934	.932	.015
.8	10¢	.732	.878	.947	.951	.954	.059

Alternatively, the primary function of monetary incentives may be to provide *S* with feedback relevant to his hypotheses, or to act as instructions as Edwards (1961) has suggested. In either of these cases little difference in effect would be expected.

Atkinson (1962) has suggested that a stimulus element may be weakly or strongly connected to a response. A Markov model based on this assumption results in the

---

[1] This research was supported by National Institute of Mental Health Grant M-3803 and National Science Foundation Grant G-22118.

[2] The data of four *S*s in the .6-0¢ group were omitted from the analysis due to an error in recording.

following asymptotic prediction:

$$P(A_1) = \frac{\pi^3 + \pi^2(1-\pi)\varphi}{\pi^3 + (1-\pi)^3 + \pi(1-\pi)\varphi} \quad [1]$$

where φ is a ratio representing the effect of loss relative to that of gain. The predictions in the last column of Table 1 were generated by inserting the mean value of φ, computed for each of the three incentive levels, into Equation 1. The mean absolute deviation of observed from predicted is less than 1.5%. Considering the variability in this data, it must be concluded that the model adequately describes these terminal response probabilities. This model also has no trouble handling the overshooting of π, and in fact will only predict π under the assumption that the effect of gain is relatively negligible (i.e., φ approaches infinity).

REFERENCES

ATKINSON, R. C. Choice behavior and monetary payoff. In J. Criswell, H. Solomon, & P. Suppes (Eds.), *Mathematical methods in small group processes.* Stanford: Stanford Univer. Press, 1962. Ch. 4.

EDWARDS, W. Probability learning in 1,000 trials. *J. exp. Psychol.*, 1961, 62, 385–394.

(Received November 8, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 5, 522–524

A REPLICATION OF PERCEPTUAL CURIOSITY AS A FUNCTION OF STIMULUS COMPLEXITY[1]

HENRY L. MINTON
Pennsylvania State University

A replication and expansion of an experiment on perceptual curiosity was conducted with a sample of 77 male volunteers. The original study indicated that curiosity significantly increased as a function of the variables of incongruity, surprisingness, relative entropy, and absolute entropy. These findings were fully replicated; all at the .001 level of significance. An attention span task, which was added to control for the ability of perceptual speed, was found to be unrelated to the experimental task. The data also indicated that the 4 variables determining perceptual curiosity reflect a unitary variable, which can best be termed stimulus complexity.

This report presents the replication and expansion of a study by Berlyne (1957) which indicated that certain properties of visual figures—incongruity, surprisingness, relative entropy, and absolute entropy—significantly increased curiosity. In addition, the present experiment provided for the evaluation of a variable—attention span—not controlled in the original investigation. Intercorrelations among the stimulus variables also were obtained.

Method.—Male volunteers from a general psychology course at Pennsylvania State University served as *S*s.

The procedure followed that of Berlyne (1957) closely. A tachistoscope, operating with two short-decay phosphor fluorescent tubes, was connected to a finger key. Whenever the key was depressed the 5 × 7 in. card in the tachistoscope became visible for .14 sec. The *S* was seated in a darkened room in front of the tachistoscope with his right hand holding the key. The *S* was told that he could have as many exposures of each figure as he wished. The number of key presses for each figure was the measure of perceptual curiosity. The stimulus material consisted of four parts, corresponding to the variables of incongruity, surprisingness, relative entropy, and absolute entropy. (Entropy is the opposite of redundancy.) The part reflecting absolute entropy also reflected the variable of meaningful sequence. These four parts followed each other without interruption. In order to ascertain the degree of validity of perceptual curiosity it was preferable to use a constant ordering of experimental series, rather than the counterbalanced Latin square design employed by Berlyne.

After the replicated experiment was completed an attention span task was presented

---

[1] This research represents part of a dissertation submitted to the Pennsylvania State University in partial fulfillment of the requirements for the PhD degree. The author is greatly indebted to Douglas N. Jackson for his valuable advice.

TABLE 1

FRIEDMAN ANALYSIS OF VARIANCE AND t TESTS OF LEVELS OF STIMULUS COMPLEXITY FOR FIVE EXPERIMENTAL VARIABLES

Experimental Variables	Source of Data	Mean Number of Responses[a]				*df*	*CR*
Congruity		Congruous cards	Incongruous cards				
	Present sample	2.6	3.5			76	$t=10.22$[b]***
	Pilot	2.5	3.3				($p<.01$)
	Berlyne[c]	3.0	5.8				($p<.001$)
Meaningful sequence		Ordered Sequence	Random Sequence				
	Present sample	2.3	2.3			76	$t=.29$
	Pilot	2.1	2.1				
	Berlyne[c]	2.4	2.4				
Surprise		Nonsurprise cards	Surprise cards				
	Present sample	2.3	2.6			76	$t=3.74$***
	Pilot	2.2	2.7				($p<.001$)
	Berlyne[c]	2.0	2.4				($p<.01$)
Absolute entropy		Low	Middle	High			
	Present	1.4	2.3	2.9		2	$\chi_r^2=105.33$***
	Pilot	1.2	2.2	2.6			($p<.001$)
	Berlyne[c]	1.6	2.3	3.1			($p<.001$)
Relative entropy Series A		A-1	A-2[d]				
	Present sample	2.0	2.9			76	$t=3.44$***
	Pilot	1.8	3.6				($p<.01$)
	Berlyne	2.6	4.2				($p<.05$)
Relative entropy Series B		B-1	B-2	B-3	B-4[d]		
	Present sample	1.4	1.4	1.8	2.3	3	$\chi_r^2=31.34$***
	Pilot	1.4	1.5	1.8	2.5		($p<.01$)
	Berlyne	1.2	1.6	1.8	3.2		($p<.01$)
Relative entropy Series C		C-1	C-2	C-3	C-4[d]		
	Present sample	1.8	2.5	2.3	4.4	3	$\chi_r^2=103.07$***
	Pilot	1.8	2.5	2.3	5.2		($p<.001$)
	Berlyne	1.7	2.6	2.8	3.3		($p<.001$)

[a] Each of the categories listed below corresponds to the breakdown of response categories.

[b] All of the t tests are two-tailed.

[c] Berlyne used a Latin square analysis. In all other cases the same statistical analysis was used as in the present study.

[d] The numbers vary in the direction of Relative Entropy.

*** $p<.001$.

to each *S*. This task was included as a control for the ability of perceptual speed, which was considered to be unrelated to the experimental task. Six cards were tachistoscopically presented in the same time interval (.14 sec.) as the experimental cards. A particular number of small black dots was scattered on each card, the number of dots being 5, 6, 7, 8, 10, and 12. The presentation of the six cards was in random order. The *S*'s task was to report how many dots he saw on each card. Only one trial per card was permitted. The six cards used in the attention span task were chosen so as to represent an adequate range of difficulty.

Results and discussion.—For each of the variables of incongruity, meaningful sequence, absolute entropy, surprisingness, and relative entropy, the mean number of responses per response category was compared. The response categories for each of the variables correspond to homogeneous levels along the relevant dimension, such as surprise cards and nonsurprise cards, or ordered sequence and random sequence. (The complete list is indicated in Table 1.) In those cases where there were only two response categories per variable t tests for dependent data were performed. In those cases where there were three or four response categories per variable Freidman (ranking) analyses of variance were conducted. These results for the present sample of 77 are shown in Table 1. Also indicated are the means from Berlyne's data ($N = 16$) and the data of a pilot study conducted by this investigator ($N = 33$). (The pilot study utilized the same procedure as the present study, but did not include the attention span task.)

The analyses of the means in the present sample replicate in full the findings of the original study and of the pilot study. Six out of the seven mean comparisons show significant differences; all at the .001 level of significance for the present sample. The one mean comparison in which no significant differences are indicated is with respect to the variable of meaningful sequence. Berlyne (1957) pointed out that this was not one of the variables central to the theory under test, and therefore this variable can be considered as of lesser relevance. The results of the other four variables indicate that they are all important factors in perceptual curiosity.

Intercorrelations among the response categories across variables were very high (the majority above .70). This would seem to indicate that an individual's pattern of responses on one of the stimulus variables is highly predictive of his pattern on another, or in other words the existing breakdown of stimulus variables are all reflecting a unitary variable. This variable can best be termed stimulus complexity, since Berlyne (1957) considers absolute—and relative—entropy as aspects of stimulus complexity and incongruity and surprisingness as closely related to stimulus complexity. In comparing the intercorrelations among response categories with the mean comparisons it appears that the variable of stimulus complexity is both significant and a consistent one in terms of levels of variation.

Intercorrelations between the attention span task and each of the response categories were also obtained. None of these intercorrelations were found to be significant at the 5% level, indicating that the attention span task is unrelated to the experimental task.

The results support the findings of Berlyne, as well as showing that the ability of perceptual speed is unrelated to the experimental task. The data also suggest that the four variables determining perceptual curiosity can be subsumed under the variable of stimulus complexity. With respect to the validity of the curiosity response, there is as yet no published evidence that individuals who press buttons are curious in other situations.

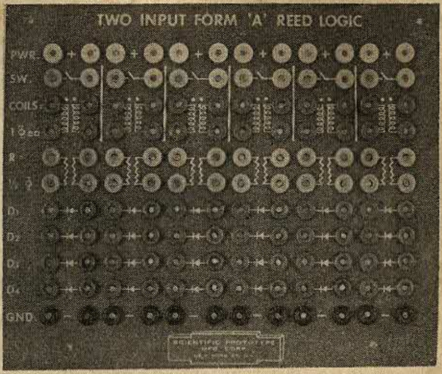
REFERENCE

Berlyne, D. E. Conflict and information-theory variables as determinants of human perceptual curiosity. *J. exp. Psychol.*, 1957, **53**, 399–404.

(Received December 3, 1962)

Scientific Prototype proudly announces . . .

THE FIRST COMPLETELY INTEGRATED REED LOGIC SYSTEM DESIGNED FOR USE IN GENERAL PURPOSE LABORATORY SYSTEMS

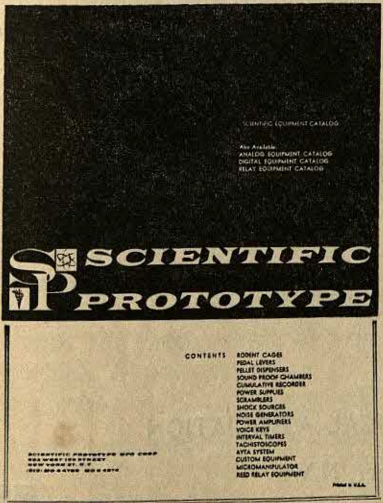


Here is a major innovation in logic and control for use in operant-conditioning programming and measurement. Here is *Reed Relay Logic,* offering all these many advantages in control systems.

- **Low cost. Universal Application.**
- **Long operating life. Measured in hundred millions of operations.**
- **Moderate-speed switching—better than 100 cps.**
- **Environmental protection with hermetically sealed switch capsules.**
- **Low power consumption.**
- **Insensitive to electrical noise pickup, even from heavy inductive loads.**
- **Maximum durability with minimum acoustic noise.**

Each reed logic unit contains six identical logic elements, including auxiliary resistors and diodes. Each reed panel occupies half a standard rack in width and clips onto standard power rods. Scientific Prototype Reed Relays offer the ultimate in flexibility. Use them individually—as gates (*and, or, nor, exclusive or*), as controlled-action switches, as flip-flops or latching relays. Use them in groups—as binary dividers or ring counters. Used in hybrid equipment with conventional timers, steppers, counters, etc., they improve performance and extend operating life.

For a detailed discussion of the characteristics and applications of this new research tool, write for our monograph "Digital Techniques" and a copy of our soon-off-the-press Reed Relay Catalog.



SCIENTIFIC EQUIPMENT CATALOG

Also Available:
ANALOG EQUIPMENT CATALOG
DIGITAL EQUIPMENT CATALOG
RELAY EQUIPMENT CATALOG

SCIENTIFIC PROTOTYPE

CONTENTS
RODENT CAGES
PEDAL LEVERS
PELLET DISPENSERS
SOUND PROOF CHAMBERS
CUMULATIVE RECORDER
POWER SUPPLIES
SCRAMBLERS
SHOCK SOURCES
NOISE GENERATORS
POWER AMPLIFIERS
VOICE KEYS
INTERVAL TIMERS
TACHISTOSCOPES
AVTA SYSTEM
CUSTOM EQUIPMENT
MICROMANIPULATOR
REED RELAY EQUIPMENT

Printed in U.S.A.

NEW 1963 CATALOG NOW AVAILABLE

Brand new: latest models, prices, information on Rodent Cages, Pedal Levers, Pellet Dispensers, Sound Proof Chambers, Cumulative Recorders, Power Supplies, Scramblers, Shock Sources, Noise Generators, Power Amplifiers, Voice Keys, Interval Timers, Tachistoscopes, Vocal Analyzers, Micromanipulators—everything.

WRITE FOR YOUR COPY TODAY!

SCIENTIFIC PROTOTYPE MFG. CORP.

623 W. 129 ST., NEW YORK 27, N. Y. (212) MO6-4789

the monthly review journal
of the
American Psychological Association

CONTEMPORARY PSYCHOLOGY

reviews

criticism

opinion

on

the latest books, films,

and other instructional media

lists of new books **letters-to-the-editor**

$1.00 a copy $10.00 a year
($10.50 foreign)

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333-16th Street, N.W.
Washington, D.C. 20036

Journal of
Experimental Psychology

VOL. 66, NO. 6 DECEMBER 1963

INTERPOLATED RECALL IN SHORT-TERM MEMORY[1]

BENNET B. MURDOCK, JR.

University of Vermont

3 experiments were conducted to determine the effect of interpolated recall on short-term memory for individual paired associates. Lists of 6 pairs were presented once, then 3 pairs (Experiment I) or all 6 pairs (Experiment II) were tested for recall. Recall tests were self-paced in Experiments I and II but paced at either a fast or slow rate in Experiment III. The results showed that recall deteriorated as number of interpolated recalls increased, but the effect was limited to 1 or 2 interpolations for pairs near the end of the list. Thereafter, the retention curve appeared to stay level at a value considerably greater than 0 over the range of retention intervals studied. The results of these and other experiments suggest that the interference generated by interpolated recall may be a rather common effect in short-term memory.

In a recent study (Murdock, 1963b), it was found that marked changes in the associative strength of individual A-B pairs occurred over a rather short period of time. Immediately following the presentation of a pair, recall probability was quite high (.90 or above). Then, during the retention interval, recall probability fell to about .60 after 2 sec., to about .30 after 4 sec., then leveled off at approximately .25 over the range of 6–10 sec. The retention interval was filled with the presentation of other pairs which *S* was trying to learn. These group results are reproducible with individual *S*s, and latencies of correct responses show comparable changes over the same retention interval (see Murdock, 1963b, Exp. IV). Finally, "forward" and "backward" associations are almost indistinguishable; for an A-B pair, recall of A given B is about as good as recall of B given A (Murdock, 1962).

The present report is a study of the effect of interpolated recall on short-term memory for individual paired associates. In general, the procedure used in the previous experiments was used here. Basically, all that we did previously was to present a list of A-B pairs once, then test for recall of one specific pair. Thus, the retention interval was filled with the presentation of "irrelevant" pairs which *S* was trying to learn and remember. Here, to study interpolated recall, more than one pair was tested for retention. In Exp. I, six pairs were presented and three pairs were tested for recall. In Exp. II, six pairs were presented and all six were tested for recall. In Exp. I and II the recall tests were self-paced; in Exp. III they were not.

---

[1] This work was supported by a research grant, M-3330, from the National Institutes of Health. Gordon Wright, Mary Jane Edwards, Howard Maltz, and Ted Edgecomb helped in testing *S*s and analyzing the data.

EXPERIMENT I

Method

If three pairs out of six are tested for recall, there are 120 different permutations. That is, the first pair tested can be one of six, the second pair one of five, and the third pair one of four possible alternatives. All 120 permutations were used twice, thus requiring 240 different lists. The experiment was conducted over four sessions with 60 lists per session. The order of presentation of the different permutations was randomized, except that each of the 120 different orders occurred once in the first two sessions and once in the second two sessions. With this design, then, the individual pairs were tested for recall in all possible orders, and the effects of interpolated recall can be determined by comparing recall of those pairs tested first, those pairs tested second, and those pairs tested third.

The pairs were composed of words selected from approximately the 4,000 most common English words (Thorndike & Lorge, 1944, G count of 20 or above). These words were randomly combined into pairs. Each pair was written on an index card. In testing, *E* would present the individual pairs to *S* at a rate of 2 sec/pair. As soon as the last pair had been presented *S* would then try to recall three of the pairs in a predetermined order. The A members of the three A-B pairs tested were also written on individual index cards; these served as the cues for recall. The recall test was self-paced; *S* went through the three cards at his own rate, and could spend as much or as little time as he wished trying to recall each of the three pairs. However, once *S* had gone from one card to the next he could not look back or change his previous answer.

All *S*s were tested individually. There were 20 *S*s, students of both sexes from the introductory psychology course who were fulfilling a course requirement.

Results

The main results are shown in Table 1, which gives recall probability as a function of serial position and order of testing. The first column gives the values for the first pair tested, and it agrees fairly well with our previous results (Murdock, 1963b). Comparison of the first and second column shows the effect of one interpolated recall, and the third column shows recall probability after two interpolated recalls.

An analysis of variance of number of correct recalls showed that the two main variables and their interaction were all highly significant; $F\ (5, 95) = 30.9$ for serial position, $F\ (2, 38) = 71.7$ for order of testing, and $F\ (10, 190) = 50.6$ for interaction, in all cases $p < .001$. Overall, there was a recency effect for serial position (the shorter the retention interval the better the recall) and a primacy effect for order of testing (interpolated recalls interfered with retention). However, the interaction term was highly significant. As can be seen from Table 1, one interpolated recall had the greatest effect on the pair in Serial Position 6, progressively less effect on pairs in Serial Positions 5 and 4, and, for all practical purposes, no effect on the remaining pairs. Sign tests (McNemar, 1962) showed that

TABLE 1

RECALL PROBABILITY FOR EXP. I

Serial Position	Order of Testing			*M*
	1	2	3	
1	.266	.315	.237	.273
2	.209	.312	.309	.277
3	.196	.240	.240	.225
4	.338	.276	.276	.297
5	.599	.340	.280	.406
6	.864	.409	.285	.515
M	.412	.315	.271	

the effect of one interpolated recall was significant at the .001, .02, and .05 levels for Serial Positions 6, 5, and 4, respectively. That is, significantly more *S*s had better recall for the first pair tested than for the second pair tested at these three serial positions. The second interpolated recall was significant by sign tests at the .001 and .02 levels for Serial Positions 6 and 5, respectively. Otherwise, there were no significant differences between adjacent cells in the same row.

In an attempt to describe the results more exactly, the relationship between recall probability and number of interpolated recalls was treated as a continuous function despite the relatively few data points available. As a first approximation, $-dp/dn$ is proportional to $(p - .25)$ where p is recall probability and n the number of interpolated recalls. The value of .25 is the asymptotic value of p, estimated from the values in Table 1 and verified with semilog plots by a method described in Murdock and Cook (1960). Thus, recall probability decreased exponentially as a function of n. Estimated rate constants for Serial Positions 6, 5, and 4 were of the order of 1.4; thus, each interpolated recall decreased $(p - .25)$ by a factor of 4. Stated simply, the effect of interpolated recall is a function of p. When p is well above asymptote one interpolation will markedly affect retention; when p is near asymptote one interpolation will have little or no effect on retention. Note that the asymptote was about .25, not .00; there was still a range left in which the effect of interpolated recall could have been manifest.

Up to this point only the number of interpolated recalls has been considered; does it matter which recall is interpolated? Analyses of the number of correct recalls of the second pair tested showed that only for the last pair presented (i.e., Serial Position 6) did the serial position of the pair tested first have a clear-cut effect. In this one case the effect of which pair was tested first was highly significant; $F(4, 76) = 10.20, p < .001$. Recall of 6-2 (i.e., Serial Position 6 tested second) was better following 5-1 than it was after 1-1, 2-1, 3-1, or 4-1. A possible explanation for this finding will be suggested in the Discussion section.

EXPERIMENT II

In Exp. I there were never more than two interpolated recalls. In Exp. II, on each trial six pairs were presented and all six were tested for recall. Thus, the number of interpolated recalls was extended from two to five. Also, some of the experimental conditions were changed to assess the generality of the results of Exp. I.

Method

If all possible permutations (i.e., ${}_6P_6$) had been used, then one replication would have required 720 different lists. This was manifestly impossible; therefore, a different counterbalancing scheme was used. For every six lists, each serial position occurred once in each of the six testing orders. Thus, a 6 × 6 Latin square was used to determine the order of testing within one replication of six lists. (As a result, each serial position was not tested equally often after every other serial position. However, the results of Exp. I suggest that the only time that this will be critical is when 5-1 is followed by 6-2. In Exp. II this particular sequence simply was not used.)

There were 36 lists per session for each of two sessions. Thus, there were 12 replications of the basic design, and a different 6 × 6 Latin square was used for each replication. There were three *E*s, each of whom tested one-third of the *S*s. Each *E* constructed his own pairs. Although the same order of presentation was used for all *S*s, the fact that there were three different sets of lists means that, in effect, idiosyncratic

TABLE 2

RECALL PROBABILITY FOR EXP. II

Serial Position	Order of Testing						*M*
	1	2	3	4	5	6	
1	.327	.301	.336	.356	.387	.372	.347
2	.346	.253	.307	.370	.233	.333	.307
3	.299	.353	.335	.293	.281	.282	.307
4	.329	.267	.318	.258	.364	.273	.301
5	.571	.367	.346	.327	.336	.295	.374
6	.755	.421	.387	.352	.384	.372	.445
M	.438	.327	.338	.326	.331	.321	

effects due to differences in item difficulty should be fairly well balanced out. There were 54 *S*s, all from the same population as Exp. I.

In other respects, the procedure was like that of Exp. I. All *S*s were tested individually, the stimulus material was written on index cards, pairs were presented at a 2-sec. rate, and the recall tests were self-paced.

Results

There were significant differences among *E*s; the results of the 18 *S*s tested by one *E* were somewhat lower than the 18 *S*s tested by the other two *E*s. Therefore, separate analyses of variance were performed on the three sets of data. The two main effects and the interaction were significant at the .001 level in each case, and the direction of the differences was the same throughout. Therefore, only the combined results will be presented.

The combined results are shown in Table 2. As a comparison with Table 1 will show, recall probability for 6-1 was somewhat lower in Exp. II than in Exp. I. Otherwise, the results were rather similar. An analysis of variance of the combined results showed that, as before, all effects were significant at the .001 level; for serial position $F\,(5, 265) = 27.0$, for order of testing $F\,(5, 265) = 32.9$, and for interaction $F\,(25, 1{,}325) = 17.6$. Again, there was a recency effect for serial position and a primacy effect for order of testing.

Perhaps the main conclusion to be drawn from Exp. II is that increasing the number of interpolated recalls from two up to a maximum of five had no apparent effect whatever. That is, mean recall probabilities were .338, .326, .331, and .321 for two, three, four, and five interpolated recalls, respectively. Even the effect of one or two interpolated recalls was negligible at the early serial positions. As in Exp. I, $-dp/dn$ was proportional to p, and once p reached a value of about .30 (±.05) no further changes occurred regardless of serial position or number of interpolated recalls.

In all there was a total of 23,328 responses (72 lists, 6 pairs per list,

TABLE 3

ANALYSIS OF RESPONSES FOR EXP. II

	*S*s		
	Bottom 25%	Middle 50%	Top 25%
Correct	.222	.339	.489
Omissions	.678	.546	.413
Errors	.101	.115	.099

and 54 *S*s). An analysis of these responses is shown in Table 3; *S*s were divided into quartiles on the basis of total number correct. As can be seen, the proportion of errors (any incorrect response) was essentially constant over *S*s; however, as correct responses increased, omissions decreased. Although not presented in detail here, a further analysis showed that approximately half of the total errors were intralist intrusions with B members more frequent than A members by a ratio of 2.2 to 1. When an intralist intrusion was a response from a prior pair it tended to immediately precede the correct response (e.g., the B member from Serial Position 4 instead of the correct B member from Serial Position 5). When an intralist B intrusion came from a subsequent pair, it was most often from Serial Position 5.

For practice effects, there was little difference between the first and second session; mean recall probabilities for all *S*s were .339 and .355, respectively. There was no consistent change over the six replications of the first session, but there was an apparent warm up (significant at the .001 level) after the first replication on the second session. Recall probabilities were .287, .389, .388, .348, .334, and .383 for Replications 1–6 in Session 2. As a measure of reliability, the correlation between total number correct on the two sessions was .86 which, with 54 *S*s, was significant at well beyond the .01 level.

EXPERIMENT III

In Exp. I and II the recall trials were self-paced. In Exp. III they were paced by *E* at one of two presentation rates. The time available for recall was either 2 sec/pair or 8 sec/pair.

Method

All lists were six pairs long, and three of the six pairs were tested for recall. In all there were 240 lists, 60 lists on each of four sessions. On the recall tests the cards with the A members of each pair were shown to *S* by *E* at a rate of either 2 sec/card or 8 sec/card. Only one rate was used after any given list, and the two rates alternated in an ABBA order on successive trials. Thus, in effect there were three variables: serial position, order of testing, and recall interval. Each *S* served as his own control and was equally often under all conditions.

The duration of the recall interval was strictly observed throughout the entire experiment. In effect, what happened was that, after each list, *E* presented the A members of the three pairs at a rate of either 2 sec/card or 8 sec/card. At the fast rate *S* had to make the correct response within the allotted time, and it was made quite clear that the response would be considered wrong unless it occurred within the 2-sec. interval. At the slow rate *S* had to wait the full 8 sec. regardless of his performance or his wishes. There was one exception to the above rules: *S*s were always given unlimited recall time for the third pair tested at the fast rate. This procedure was adopted so as to have at least one valid comparison between the two rates. That is, this made it possible to determine the effects of elapsed time with number of interpolated recalls held constant.

In other respects the procedure was much like that of Exp. I. The same A-B pairs were used, and all 120 permutations of ${}_3P_6$ were used twice, once on the first two sessions of 60 lists each and once on the second two sessions of 60 lists each. Each of the 18 combinations of Serial Position × Order of Testing occurred equally often at the 2-sec. rate and at the 8-sec. rate. However, no attempt was made to insure that each of the 120 permutations occurred once at each of the two presentation rates. There were 20 *S*s, all from the same group as before. All *S*s were tested individually.

Results

The main results of Exp. III are presented in Table 4. In general, the results for the 8-sec. recall interval do not seem to be too different from the results of the first two experiments. At the 2-sec. interval, how-

TABLE 4

RECALL PROBABILITY FOR EXP. III

Serial Position	Order of Testing					
	2-Sec. Interval			8-Sec. Interval		
	1	2	3	1	2	3
1	.318	.308	.238	.328	.354	.264
2	.162	.312	.285	.258	.398	.340
3	.215	.212	.225	.201	.318	.306
4	.301	.244	.386	.274	.333	.320
5	.458	.388	.335	.572	.412	.425
6	.588	.574	.412	.692	.404	.378

ever, the recall probabilities were somewhat lower.

An analysis of variance of the number of correct recalls showed that the effects of all single variables (Serial Position, Order of Testing, and Recall Interval), two of the first-order interactions (Serial Position × Order of Testing and Serial Position × Recall Interval), and the second-order interaction (Serial Position × Order of Testing × Recall Interval) were all significant at least at the .01 level. The one exception was Order of Testing × Recall Interval; here $F < 1.00$. Thus, the effect of interpolated recalls did not vary with the time available for recall; comparable decrements with one and two interpolated recalls were obtained at both the 2-sec. and the 8-sec. pacing.

The significant interaction between Serial Position and Order of Testing shows that, as in the two previous experiments, the effect of interpolated recall varied as a function of the recall probability. Actually, when summed over the 2-sec. and the 8-sec. recall interval, the exponential relationship again seemed not unreasonable. Thus, insofar as one can say, the proportionality between dp/dn and p occurs with paced recall as well as with unpaced recall. However, the second-order interaction was significant and, as can be seen in Table 4, the interaction of Serial Position and Order of Testing appears to be less pronounced with a 2-sec. recall interval than with an 8-sec. recall interval.

Unlimited recall time was allowed for the third pair tested. When summed over all six serial positions recall probabilities were .317 for the 2-sec. condition and .338 for the 8-sec. condition. Thus, even though considerably more time had elapsed in the latter condition, recall was slightly better. Of course, with two interpolated recalls, initial recall probability would have declined by a factor of 16; therefore, it is not too surprising that differences in amount of elapsed time had little apparent effect.

Perhaps the most unexpected finding in the experiment was the nature of the differences between the 2-sec. and the 8-sec. recall interval for the first pair tested. In a previous study (Murdock, 1963b) there was a high correlation between accuracy and latency. Moreover, when only one pair per list was tested for recall, 97.3% of the correct responses for Serial Position 6 had latencies less than 2 sec., while only 56.2% of the correct responses for Serial Positions 1–3 had latencies less than 2 sec.

Therefore, one would expect that the effect of curtailing the recall interval would interact with serial position. It did, but in the wrong way; the difference between the two recall intervals was .101 for Serial Position 6 but only .031 for Serial Positions 1–3. By sign tests, the former difference was clearly significant, but the latter difference just made the .05 level, one-tailed. We have no explanation for why (*a*) the difference for Serial Position 6 was so large; or (*b*) the difference for Serial Positions 1–3 was so small.

DISCUSSION

The results of these three experiments leave little doubt that interpolated recall has an effect on short-term memory for paired associates. Not only were the effects highly significant in all three cases, but the overall pattern of results was very similar throughout. Further, it is quite clear that interpolated recall will have a marked effect when recall probability is high but little or no effect when recall probability is low.

There are really two aspects to this effect of interpolated recall. One aspect is the marked effect one or two recalls will have on the last pair or two in the list. The other aspect is the complete lack of effect interpolated recall has throughout much of the range studied here. Recall probability quickly approaches an asymptote of about .25 or .30, and it seems to stay there even when some of the early pairs in the list are tested near the end. Actually, a fair share of the data in Tables 1, 2, and 4 simply demonstrates that recall probability does seem to stay constant over a fairly wide range.

In Exp. I, recall of 6-2 was better if preceded by 5-1 than if preceded by pairs from other serial positions. Time factors are probably not involved here since the same effect occurred in Exp. III with paced recall. One possible explanation is as follows.[2] Perhaps on a recall test *S* has available a limited pool of responses from which to select the correct response. If he were able to eliminate one response from this pool, then he would have a better chance of recalling the pair in question. If the pair from Serial Position 5 had been tested first he would have had at least a fair chance of being correct on his first recall. Then the response from Serial Position 5, having just occurred, could easily be eliminated from the pool. If pairs from other serial positions had been tested first they would be less likely to be correct, hence could not be so easily eliminated.

Such an explanation clearly requires that *S* recall 6-2 more often given that 5-1 had been correct than given an incorrect 5-1. This in fact happened: the conditional probabilities were .73 and .49, respectively. By a test for independent proportions this difference was significant at the .01 level. The same effect occurred in Exp. III, but there the two conditional probabilities (.69 and .60, respectively) did not differ significantly. A mechanism of this sort has been proposed to account for certain aspects of recognition (Murdock, 1963a); perhaps the same effect can occur in recall.

Finally, there are numerous other studies which have shown the effect of interpolated recall. Results similar to those reported here have been found by Brown (1954), Lawrence and Laberge (1956), Broadbent (1958), Anderson (1960), Sperling (1960), and Mackworth (1962). Also, in one study Conrad (1960) found that merely interpolating the known digit "0" prior to the recall of a list of eight digits markedly decreased the number of lists correctly recalled. Lastly, Sanders (1961) has shown that immediate recall can even interfere with subsequent recall of the same material. Taken together, these results suggest that the interference generated by interpolated recall may be rather common. Perhaps RI and PI are not the only types of interference in short-term memory.

---

[2] We are indebted to Harold Dale for several stimulating discussions of these ideas.

REFERENCES

Anderson, N. S. Post-stimulus cuing in immediate memory. *J. exp. Psychol.*, 1960, **60**, 216–221.

Broadbent, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

Brown, J. The nature of set-to-learn and of intra-material interference in immediate memory. *Quart. J. exp. Psychol.*, 1954, **6**, 141–148.

Conrad, R. Very brief delay of immediate recall. *Quart. J. exp. Psychol.*, 1960, **12**, 45–47.

Lawrence, D. H., & LaBerge, D. L. Relationship between recognition accuracy and order of reporting stimulus dimensions. *J. exp. Psychol.*, 1956, **51**, 12–18.

McNemar, Q. *Psychological statistics* (3rd ed.) New York: Wiley, 1962.

Mackworth, J. F. The effect of the response upon the immediate memory span. *Canad. J. Psychol.*, 1962, **16**, 120–127.

Murdock, B. B., Jr. Direction of recall in short-term memory. *J. verbal Learn. verbal Behav.*, 1962, **1**, 119–124.

Murdock, B. B., Jr. An analysis of the recognition process. In C. N. Cofer (Ed.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. (a)

Murdock, B. B., Jr. Short-term retention of single paired associates. *J. exp. Psychol.*, 1963, **65**, 433–443. (b)

Murdock, B. B., Jr., & Cook, C. D. On fitting the exponential. *Psychol. Rep.*, 1960, **6**, 63–69.

Sanders, A. F. Rehearsal and recall in immediate memory. *Ergonomics*, 1961, **4**, 25–34.

Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received December 26, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 533–541

EFFECT OF SPACING AND RANGE OF STIMULI ON MAGNITUDE ESTIMATION JUDGMENTS[1]

PANNA LAL PRADHAN AND PAUL J. HOFFMAN
College of Education, Kathmandu, Nepal *Oregon Research Institute, Eugene*

This is an investigation of (a) the extent to which individual psychophysical functions are power functions, and (b) the effect of spacing and range of stimuli on the psychophysical function. The 6 experimental conditions consisted of 2 ranges and 5 spacings of the stimuli. A balanced Latin square design was followed, involving Os, conditions, and order of the sessions. In each session O judged the heaviness of each of 9 weights 20 times by the method of magnitude estimation. Individual psychophysical functions were found not to follow Stevens' law although averaging over Os does yield a power function. Stevens' power function thus seems to be an artifact of grouping. Significant differences emerged for the various conditions of spacing, but these differences cannot be unambiguously attributed to the effect of spacing, space-individual interaction, or the ordering of the sessions.

The primary purpose of this study is to investigate the effect of spacing and range of stimuli on psychophysical functions obtained by the method of magnitude estimation (Stevens, 1958, p. 193).

According to Stevens, the psychophysical function obtained from magnitude judgments is a power function of the form

$$R = aS^n \qquad [1]$$

in which R is the magnitude of the stimulus on the psychological scale, S is the magnitude of the stimulus on the physical scale; both R and S are measured on ratio scales; n is a parameter peculiar to the psychological attribute; and a is a constant determined by the arbitrary selection of the unit of measurement. (For alternative formulations of the power law, see Ekman, 1958; Ekman & Kunnapas, 1961; Fagot, 1962; Galanter & Messick, 1961; McGill, 1960; and Stevens, 1959, 1960a, 1960b.)

Taking logarithms of both sides of Equation 1, we have:

$$\log R = \log a + n \log S \qquad [2]$$

Thus, $\log R$ is a linear function of $\log S$ with slope n, and the test of linearity of regression of $\log R$ on $\log S$ is an appropriate test of the adequacy of the psychophysical power function; i.e., if the regression is not significantly different from linearity, the power function does not hold.

The conclusion arrived at by S. S. Stevens is based on the data (medians) obtained from *groups* of *O*s. But Sidman (1952) and Estes (1956) have shown that the relationship obtained from the average of the data over a group of *O*s does not necessarily specify the form of the relationship for the individual *O*s. This would very likely hold also for medians,

---

[1] This study was supported in part by National Science Foundation Grant G19210 to the University of Oregon under the direction of Robert F. Fagot and in part by Public Health Service Grant M-2097 from the National Institute of Mental Health to Oregon Research Institute under the direction of Paul J. Hoffman.

Acknowledgment is made of the computer facilities available at the Statistical Laboratory, University of Oregon, and at Western Data Processing Center, University of California, Los Angeles, which were used for this study.

depending only upon the symmetry of the distributions, although Jones and Marcus (1961) are of the opinion that log mdn. R is a reasonable approximation of mdn. log R. Estes (1956) has shown that the relationship obtained on the group data is of the same form as that obtained on the individual data only if the function has a form such that all the second and the higher-order partial derivatives of the function with respect to each of the parameters are zero. If, on the other hand, the function in the original form does not satisfy the above criterion but some transformation of it does, then the relationship obtained with the data transformed to this form and then averaged has the same form as the individual function under the same transformation. Applying this criterion to the power function, we have:

$$\frac{\delta^2 R}{a^2} = 0 \quad \text{and} \quad \frac{\delta^2 R}{n^2} \neq 0$$

but

$$\frac{\delta^2 (\log R)}{\delta (\log a)^2} = 0 \quad \text{and} \quad \frac{\delta^2 (\log R)}{\delta n^2} = 0$$

It therefore becomes obvious that if the individual functions are power functions, the group function is a power function provided the *geometric* mean is used to average the data. However, if the group function is a power function, there is no constraint that the individual functions be power functions, since both the median and the mean are invariant with respect to a wide variety of transformations of R.

A few studies in which the individual functions have been reported suggest that the power function might not be the proper psychophysical functions for the individual *O*s (cf., Richardson & Ross, 1930). In the paper by Jones and Marcus referred to previously, individual functions were modified to include a constant characteristic of the individual S. If C_i is an appropriate constant for S_i (for the ith individual), these authors express the psychophysical function as

$$R = aS^{nC_i}$$

and then suggest that averaging over individuals is appropriate if a logarithmic transform is first used or if the geometric mean is taken. Jones and Marcus (1961) conclude that in spite of seemingly good success in fitting the data by their method, ". . . it is not possible to be certain . . . that there was not significant curvilinearity in the data." The appropriateness of the power function for individuals therefore remains an open question.

Few studies have explored the effect of spacing and range of stimuli on the psychophysical function obtained by the method of magnitude estimation. J. C. Stevens (S. S. Stevens, 1956) has found that the psychophysical functions (loudness) obtained by this method with two different spacings of stimuli—log and sone—were almost identical. On the basis of this study, S. S. Stevens (1956) concluded that "the spacing of the variable stimuli seems to play only a negligible part in determining the estimate made in these experiments [p. 13]." But such conclusions appear unwarranted, since the two psychophysical curves were group curves. The individual-spacing interactions would tend to cancel out in the process of averaging.

A study by Kunnapas (1960) illustrates the same shortcomings.

If the power law holds, it must hold not only for sets of stimuli with differing means, but also for sets of stimuli with differing ranges, and it should hold for individuals as well as for group data. The present study is addressed to these questions.

METHOD

Apparatus.—A classroom chair with extended right arm forming a desk constituted the main body of the apparatus. An 18 × 24 in. plywood board was attached to the desk. There were two holes in the board, 3.9 in. apart. A flanged glass tube was inserted through each hole and a cord passed through each tube. A tab was attached to the upper end and a hook to the lower end of the cord. Seven and one-half inches of the cord were outside the tube. The chair was placed on top of a table. The *O* sat in the chair, and *E* hung the desired weights by the hooks. The *O* estimated the heaviness of the weights by pulling the cords as many times as was necessary for him to reach a decision. The arrangement of the apparatus concealed the weights being used from the view of the normally seated *O*.

Subjects.—The *O*s were six undergraduate students from the University of Oregon, three males and three females. Each was approximately 19 yr. old.

Design.—There were six different experimental conditions. The conditions consisted of two different ranges and five different spacings of the stimuli. In each condition there were nine different variable stimuli (S) and one standard. The standard was the same in each condition and weighed 300 gm. The lightest and the heaviest variable stimuli were 10 and 506 gm., respectively, in the longer range and 240 and 360 gm. in the shorter range. The other seven variable stimuli were chosen to the nearest gram to obtain an equal spacing in x when $x = f(S)$. All five spacings were used in the longer range, and only one spacing was used in the shorter range. The following are the six conditions and the principles by which the stimuli were chosen. (*a*) Longer range: Cond. I,[2] logarithmic spacing, $x = \log S$; Cond. II, exponent .5 spacing, $x = S^{.5}$; Cond. III, exponent 1.0 or linear spacing, $x = S$; Cond. IV, exponent 1.5 spacing, $x = S^{1.5}$; Cond. V, exponent 2.0 spacing, $x = S^{2.0}$. (*b*) Shorter range: Cond. VI, linear spacing, $x = S$.

Each of the six *O*s was run through eight sessions of the experiment. The first six sessions consisted of the six conditions—one condition per session. In the seventh and eighth sessions *O* repeated the work done in the second and first sessions, respectively. A balanced Latin square design was used for the first six sessions, involving *O*s, conditions, and order, and balanced for condition and order.

Procedure.—One or 2 days before the beginning of the experiment, each *O* was called in and was acquainted with the task. Each *O* was run through a practice trial of 90 observations at that time. For this practice run, a combination of the lightest (10 gm.), the heaviest (506 gm.), and seven other weights drawn randomly from the six conditions was used. The *O* was seated in the chair. Then, *E* said:

> I am going to present you a series of weights, and your task is to estimate their heaviness. You will do this by assigning a number to each of them proportionate to its felt heaviness. (Hanging the standard by the left cord.) This we will call 100. (Hanging a weight by the right cord.) If that one is 100, how heavy will this be? Pull these cords as many times as you need, and then tell me the number that will represent (pointing to the right one) its heaviness if that one (pointing to the left one) is 100. You can use any number—decimal, fraction, or whole number—the only restriction being that the number you use should be proportional to the felt heaviness; that is, if the weight in this (right) cord feels twice as heavy as the standard, the one in the left, say 200; if 10 times, say 1,000; if half, say 50; if one-hundredth, say 1, and so on. The standard will always be there, so in every trial when I say go, you pull this cord (right), then pull that one (left), and then pull this one (right), and so on. You may do this as many times as you need, but your last pull also should be this one (right). Do not try to be consistent. Make your judgments independent of what you have done in the past. Every time, compare the given weight with the standard only, and then tell me the number proportionate to its felt heaviness.

A few more weights were given, and *O* was asked if he had any questions. Then the practice session of 90 observations (as described earlier in this section) was run. Time taken to complete this series was noted and was used to schedule the subsequent sessions.

---

[2] The following weights were used in Cond. I: 10, 16, 27, 44, 70, 116, 169, 310, and 506 gm. Due to an error in computation, 169 gm. was used instead of the 189 gm. which would have been more suitable for log spacing.

On the subsequent evenings, before the beginning of the experiment, *O* was reminded of the task; "You have to do the same kind of task that you did before. Make your judgments independent."

Altogether, 180 observations were taken in each session—20 observations per stimulus. The total number of observations was divided into two blocks of 90 observations each—10 observations per stimulus per block. The order of presentation of the stimuli in each block was random. But for all *O*s, the same order was followed in each session. An interval of 3–5 min. was given between the blocks. None of the weights being used was visible to *O*, and he was not aware of the plan of the experiment.

RESULTS

Individual data.—The power function implies that the regression of log *R* on log *S* is linear. Therefore, the test of linearity of the log *R* vs. log *S* curve is the test of the appropriateness of the power function as the psychophysical function. The conventional test of linearity of regression is based on the assumption that the variances of log *R* for different *S*s are equal. This requirement is not satisfied for the present data, as Fig. 1 illustrates. Instead, a test of linearity of regression was devised, based on the technique of estimation of simultaneous confidence intervals for several means (Dunn, 1961; Sheffé, 1953, 1958). The result of this test is presented in Table 1.

The regression of log *R* on log *S* is significantly nonlinear in most or all of the longer range conditions in the case of five *O*s. Results for one of the *O*s are plotted in Fig. 2, as illustrative of these findings. For only one *O* is the regression linear under every condition. The forms of the nonlinearity of the regression are different for different *O*s, but there is considerable intra-*O* consistency. The curves obtained in RP have steeper slope in the upper section than in the lower

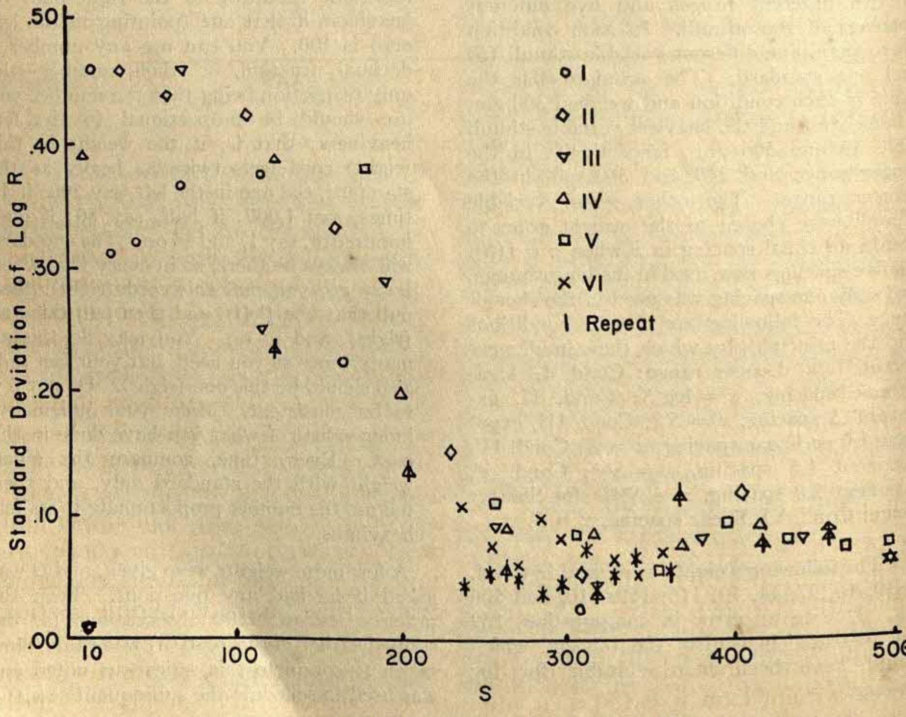


FIG. 1. Standard deviation of log *R* as a function of *S* for CB in all conditions.

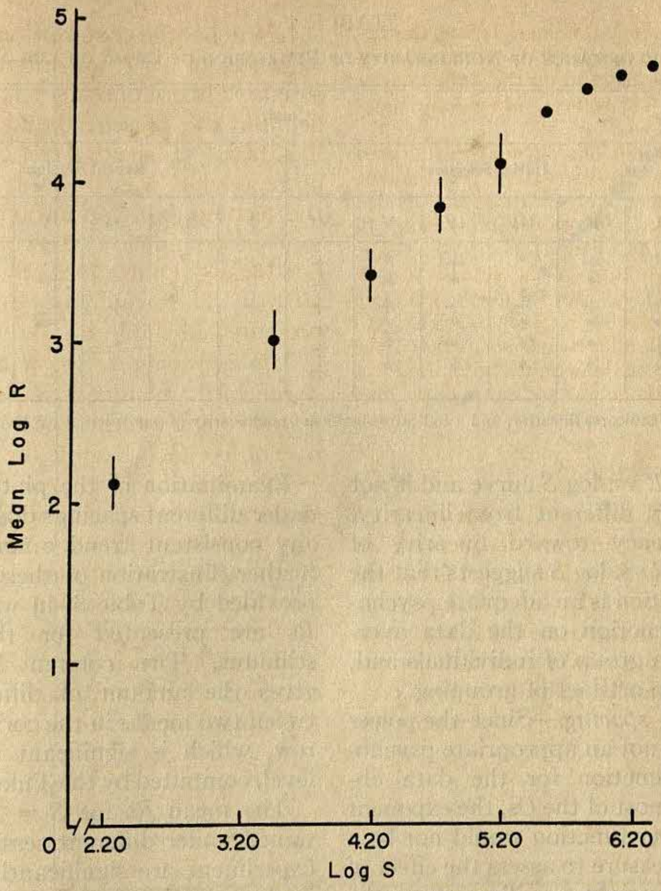


FIG. 2. Mean log R as a function of log S for CP in Cond. II. (The vertical line is the 95% confidence interval of mean log R.)

section. Those obtained on CB, CP, and GD have steeper slope in the lower section than in the upper section. Those obtained on AC are concave upwards.

Another important feature of the curves is their general slope—the first degree coefficient of regression. (If the power function were adequate, this would be the exponent, n, of the function $R = aS^n$.) These coefficients are presented in Table 2.

The interindividual differences are sufficiently large relative to intraindividual variance as to be nonchance. And since the form of the curves obtained on five *O*s is significantly nonlinear, the power function is not an adequate function to describe the psychophysical function obtained on them. Only the data obtained on GH are consistent with a power function.

Group data.—The group mean log R corresponding to the selected values of S in the five initial longer range conditions has been plotted in Fig. 3 with log S on the abscissa. The upper and the lower sections of this curve are quite straight and fall along the same straight line, and the middle section is slightly concave upwards. However, this curve is straighter than every significantly nonlinear indi-

TABLE 1

SIGNIFICANCE OF NONLINEARITY OF REGRESSION OF LOG R ON LOG S

*O*s	Cond.											
	Initial Sessions						Repeat Sessions					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI
RP	*	+	*	+	*	+		*				*
CB	+	*	*	*	*	+				*		+
GH	+	+	+	+	+	+	+		+			
CP	*	*	*	*	*	+		*		*		
GD	+	*	*	*	*	+			*		*	
AC	*	*	*	*	*	+	*					

Note.—+ indicates linearity, and * indicates significant nonlinearity of regression at .05 level.

vidual log R vs. log S curve and is not significantly different from linearity. This tendency toward linearity of group log R vs. log S suggests that the power function is an adequate psychophysical function on the data averaged over a group of individuals and, hence, is an artifact of grouping.

Effect of spacing.—Since the power function is not an appropriate psychophysical function for the data obtained on most of the *O*s, the exponent of the power function would not be a suitable measure to assess the effect of different variables. To assess the effect of spacing and range of stimuli on the psychophysical function, the mean R vs. S curves and mean *R*s for *S*s which are common in two or more conditions were used.

Examination of the plots obtained under different spacings does not show any consistent trend among *O*s. A further illustration of these results is provided by Table 3, in which mean *R*s are presented for the 10-gm. stimulus. The column headed T gives the amount of difference between two means in the corresponding row, which is significant at the .05 level, computed by the Tukey method.

The mean *R*s for $S = 10$ gm. obtained under different sessions of the experiment are significantly different in the case of five *O*s (all *O*s tested for the significance of this difference). But the difference between the mean *R*s obtained under different conditions cannot be unambiguously attributed to the difference in the spacing of the

TABLE 2

FIRST-DEGREE COEFFICIENT OF REGRESSION OF LOG R ON LOG S

*O*s	Cond.											
	Initial						Repeat					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI
RP	.87	.92	.76	.85	.80	1.75		.86				2.32
CB	1.35	1.44	1.38	1.30	1.33	1.20				1.35		.91
GH	.83	.96	.87	.87	.83	.89	.90		.84			
CP	.45	.68	.55	.42	.43	.26		.52		.44		
GD	.65	.58	.84	.78	.81	.20			.53		.59	
AC	1.91	1.95	1.98	1.94	1.66	1.75	1.55				1.50	

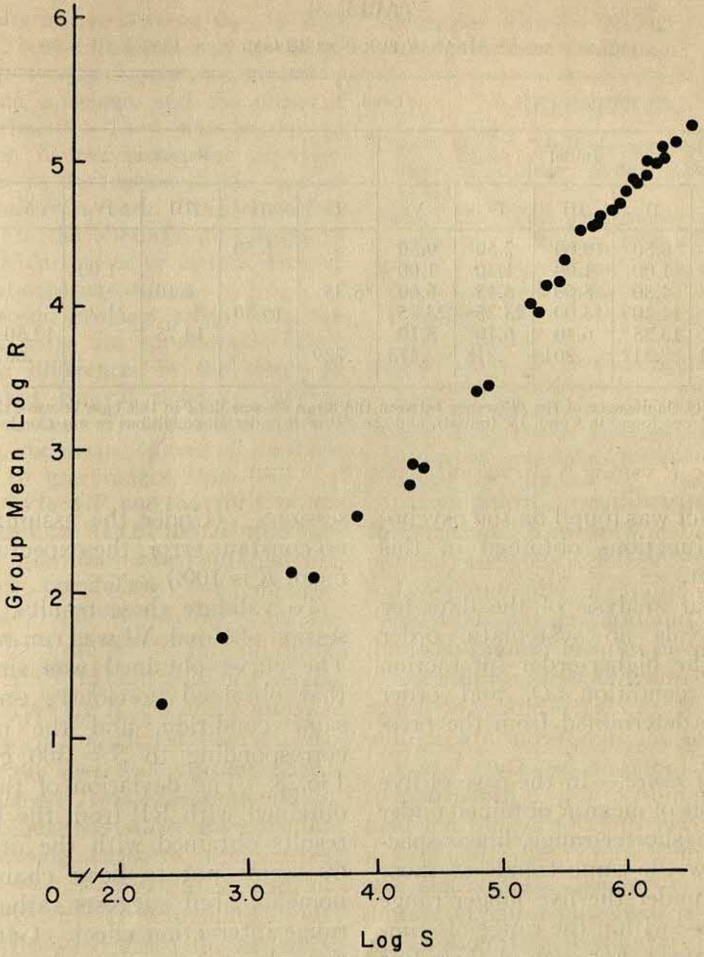


FIG. 3. Group mean log *R* as a function of log *S*.

stimuli because (*a*) there is no special trend in the mean *R*s that is peculiar to spacings of the stimuli over all *O*s (the coefficient of concordance, *W*, among *O*s was .07, not statistically significant); (*b*) in many cases the difference between the mean *R*s obtained in the two replications of the same condition is also significant; for example, GH in Cond. I, CP in Cond. II, GD in Cond. III, and AC in Cond. I and V; and (*c*) in some cases the difference between the mean *R*s in the two replications of the same condition is the largest difference; for example, CB in Cond. IV and GD in Cond. III.

The mean *R*s for $S = 506$ gm. yield similar results. The coefficient of concordance among *O*s in this case is .08, also not significant. Significant differences were found between mean *R*s for $S = 310$ gm. in the case of GH and for $S = 320$ gm. in the case of GH and CP. In each case, however, the significant difference found was between the mean *R*s obtained in two replications of the same condition. Hence it can be concluded that no systematic spacing or spacing inter-

TABLE 3

MEAN R FOR S = 10 GM.

Os	Cond.										*T*
	Initial					Repeat					
	I	II	III	IV	V	I	II	III	IV	V	
RP	6.50	6.50	10.00	7.50	9.50		6.50				2.89*
CB[a]	1.00	1.00	1.00	1.30	1.00				1.00		
GH	7.85	4.30	5.60	5.45	6.00	5.35		6.10			1.70*
CP	21.00	11.10	15.00	23.25	23.25		16.50		21.25		5.00*
GD	11.50	13.25	6.10	6.70	5.70			14.75		12.50	2.29*
AC	.224	.312	.201	.218	.475	.729				.687	.150*

[a] No test of significance of the difference between the mean *Rs* was done in this case because the only different mean R was found in Cond. IV (initial), and the *SD* of R under all conditions except Cond. IV (initial) was zero.

* $p < .05$.

action effect was found on the psychophysical functions obtained in this experiment.

A similar analysis of the data for order reveals no systematic order effect. The higher-order interaction involving condition, *O*, and order cannot be determined from the present data.

Effect of range.—In the case of five *O*s the plots of mean R obtained under Cond. VI (shorter range, linear spacing) follow the same trends as those obtained under the five longer range conditions—within the range of stimuli concerned. For each of these five *O*s the plot obtained under Cond. VI is entirely within those obtained under the longer range conditions. The plots of *SD* of log R (Fig. 1) also give the same result. Hence, in the case of these five *O*s the range of stimuli has no apparent effect on the psychophysical function.

In the case of the remaining *O* (RP), the plots of mean R obtained under Cond. VI (both initial and repeat) are steeper than the rest of the plots in that range of S. The mean R corresponding to S = 300 gm. ranges from 98.30 to 100.85 for the five *O*s. In the case of RP these values are 112 in the initial and 155 in the repeat sessions. (Under the assumption of no constant error, the expectation for mean R is 100.)

To validate these results, another session of Cond. VI was run with RP. The curve obtained was similar to that obtained previously under the same condition, and the mean R corresponding to S = 300 gm. was 136.25. The deviation of the result obtained with RP from the trend of results obtained with the other five *O*s seems not to be a chance phenomenon, but suggests rather an *O*-range interaction effect. Generalization about interactions involving *O*s are of course hazardous with this N.

DISCUSSION

The results bring forth several important problems. The first of these concerns intraindividual differences. The psychophysical curves obtained for the same *O* with different spacings of S have the same general trend. However, significant differences between the mean *R*s obtained for the same *O* and the same S under different sessions of the experiment have been found. It has been shown that these differences cannot be unambiguously attributed to the effect of spacing, spacing-individual interaction, or the order of the sessions of the experiment. Then what causes the differences?

Several alternative answers may be suggested. These differences may be due to some higher-order interaction involving individuals, spacings, and the order of the experiment. They may be due to fluctuation in the perception of magnitude, or to fluctuation in the way of using numbers, or both. It is impossible to attribute the obtained differences to any particular factor or factors without additional experimentation.

The second problem arises from the fact that there are considerable interindividual differences in the shape as well as the general slope of psychophysical curves. To take a few specific examples, the mean *R* (over all sessions) for $S = 10$ gm. ranges from .407 for AC to 18.76 for CP, and that for $S = 506$ gm. ranges from 115.82 for CP to 369.39 for AC. In this case also, several alternative possibilities may be suggested: (*a*) individuals differ in the perception of magnitude (this would imply that the perceived heaviness of $S = 10$ gm. was 46 times heavier for CP than for AC and that for $S = 506$ gm. it was 3.2 times heavier for AC than for CP); (*b*) the perceived magnitude of a certain weight is the same for all individuals, but individuals differ in their ways of using numbers—that is, the psychophysics of numbers is different for different individuals; (*c*) individuals differ both in the perception of magnitude and in the way of using numbers; (*d*) individuals are not capable of estimating perceived magnitude on a ratio scale.

The third problem arises from the fact that the psychophysical curves obtained for RP under Cond. VI were very different from those obtained under the other five conditions. Since this sort of difference was not found with any other *O*, it must be due to some factors which were not operating for the others.

The study of the problems related to the intra- and interindividual differences in the psychophysical functions and the explorations of the factors that cause these differences should be of obvious importance. It is not possible to attach meaning to the psychophysical function obtained by the method of magnitude estimation without having a reasonable solution to these problems.

REFERENCES

Dunn, O. J. Multiple comparison among means. *J. Amer. Statist. Ass.*, 1961, **56**, 54–64.

Ekman, F. Two generalized ratio scaling methods. *J. Psychol.*, 1958, **45**, 287–295.

Ekman, F., & Kunnapas, T. Scotopic brightness of monochromatic light. Report No. 100, 1961, Stockholm University, Psychological Laboratory.

Estes, W. K. The problem of inference from curves based on group data. *Psychol. Bull.*, 1956, **53**, 134–140.

Fagot, R. F. On the psychophysical law and estimation procedures for equisection and prescribed ratio scaling. Eugene: University of Oregon, 1962. (Mimeo)

Galanter, E., & Messick, S. The relation between category and magnitude scales of loudness. *Psychol. Rev.*, 1961, **68**, 363–372.

Jones, F. N., & Marcus, M. J. The subject effect in judgments of subjective magnitude. *J. exp. Psychol.*, 1961, **61**, 40–44.

Kunnapas, T. Scales for subjective distance. *Scand. J. Psychol.*, 1960, **1**, 187–192.

McGill, W. J. The slope of loudness functions. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling.* New York: Wiley, 1960.

Richardson, L. F., & Ross, J. S. Loudness and telephone current. *J. gen. Psychol.*, 1930, **3**, 288–304.

Scheffé, H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 381–400.

Scheffé, H. *The analysis of variance.* New York, Wiley, 1958.

Sidman, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, **49**, 263–269.

Stevens, S. S. The direct estimation of sensory magnitude-loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

Stevens, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, **55**, 177–196.

Stevens, S. S. Tactile vibration: Dynamics of sensory intensity. *J. exp. Psychol.*, 1959, **57**, 210–218.

Stevens, S. S. Psychophysics of sensory function. *Amer. Scientist*, 1960, **48**, 226–258. (a)

Stevens, S. S. Warmth and Cold: Dynamics of sensory intensity. *J. exp. Psychol.*, 1960, **60**, 183–192. (b)

(Received December 8, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 542–546

PERSISTENCE OF NONREINFORCED RESPONDING AS A FUNCTION OF THE DIRECTION OF A PRIOR-ORDERED INCENTIVE SHIFT [1]

MELVIN H. MARX, JO WOOD TOMBAUGH, CHARLES COLE, AND DENIS DOUGHERTY

University of Missouri

After a pretraining period in which all Ss were trained to bar press for water, 2 groups of rats were given 4 30-sec. reinforced trials and 4 30-sec. nonreinforced trials daily for 12 days. During each reinforced trial, Ss received continuous reinforcement of a sucrose solution. The concentration of the sucrose was progressively shifted either up or down over the 4 reinforced trials. The group which received the increasing incentives bar pressed reliably less than the decreasing-incentive group during the reinforced phase, but reliably more during the nonreinforced test. Also, the interaction between the terminal reinforced trial and the 1st test trial was reliable, with the increasing group pressing less during the last reinforced trial, but more during the 1st nonreinforced trial. These results support an interpretation of extinction as due to a motivational decrement rather than the terminal level of reinforced performance, and are inconsistent with several other currently popular interpretations.

This experiment was designed to test certain implications of a motivational theory of extinction (Marx, 1958, 1960; Marx & Murphy, 1961) in which a decrease in *S*'s motivation is held to be the primary factor responsible for the decrement in response that occurs during extinction.

In the present experiment motivation was manipulated by shifting the value of an incentive progressively up or down during reinforced responding. It was assumed that an animal offered progressively more desirable incentives would be more strongly motivated to respond during subsequent nonreinforced trials than an animal offered progressively less desirable incentives.

The reinforcers employed were concentrations of sucrose, made available for short periods of continuous reinforcement. Animals receiving weak concentrations of sucrose (4 and 8%) bar press for these incentives at a relatively high rate whereas animals receiving saturated concentrations (32 and 64%) bar press at a relatively low rate because of the rapid satiation effects (Collier & Willis, 1961). On the assumption that the desirability of the reinforcer varies directly with the sucrose concentrations, the group receiving the increasing order of concentration should display a relatively higher rate of responding during nonreinforced (extinction) test periods than the decreasing order group, in spite of a lower rate during acquisition.

This prediction is to be contrasted with a generalization proposed by Pubols (1960), on the basis of his review of the literature on amount of reinforcement. He suggested that nonreinforced performance can be attributed to the terminal level of reinforced performance. Utilization of an increasingly concentrated liquid incentive which results in decreasing

---

[1] This research was supported in part by United States Public Health Service Grant MH 00817-09 from the National Institutes of Health.

reinforced performance makes an experimental comparison of the two positions possible.

Method

Subjects

The *S*s were seven adult male albino and seven adult male hooded rats with no prior bar press experience or exposure to sucrose concentrations. Two additional hooded rats, one from each experimental group, were dropped from the experiment for failing to reach the bar press criterion in pretraining.

Apparatus

Eight operant conditioning boxes were used. The boxes were constructed of tempered Masonite and were 8¾ in. long × 6 in. wide × 7½ in. high. Each box contained a 6-w. light and a cylindrical stainless steel bar, 2 in. long. and ½ in. in diameter, inserted ¾ in. into the box. Depression of the bar operated a dipper, with a capacity of .1 ml. The bar could be locked from the outside to prevent depression. Bar presses were recorded on Veeder-Root counters.

Experimental Design

All *S*s were first pretrained, under water deprivation, to bar press for water. Periods of reinforcement (bar unlocked, light off) and nonreinforcement (bar locked, light on) were alternated. The *S*s were then divided into two equated groups on the basis of their bar press rates. Group I received the increasing, Group D the decreasing sucrose concentrations. Four reinforced sessions, with the increasing or decreasing concentration order, and four immediately following nonreinforced test trials were given daily for 12 days.

Procedure

Pretraining.—Prior to the beginning of the experiment, all of the animals were placed on 23-hr. water deprivation for 7 days. On the following 3 days, 23 hr. after watering, each animal received a daily 20-min. bar press training period with water reinforcement. During this phase, the light was off in each box. All animals were trained to a criterion of a total of 15 bar presses during the final 5 min. of each of the last two periods. This procedure required three training periods per animal. Four animals were not able to reach criterion during this time and received an additional 20 min. of bar press training; two of them still failed to bar press adequately and were discarded.

Discrimination training.—At the beginning of discrimination training, *S*s were placed on food as well as water deprivation. On the next 9 days, the animals were subjected to training procedures designed to make the dark cage condition an S^D for bar pressing. Water reinforcement was used. In general, the procedure consisted of gradually increasing the length of the periods when the bars were locked, with light on, and gradually decreasing the length of the periods when they were operative, with lights off. This procedure was instigated in order to prevent extinction while discrimination was being established. The length of the periods during which the bars were inoperative varied over days from 15 to 120 sec. On the final 2 of the 9 days of discrimination training, each *S* received 8 30-sec. trials with 120-sec. intertrial intervals. All bar presses within each 30-sec. trial were reinforced.

The performance of *S*s on the last 12 reinforced trials during the final 2 days of discrimination was used to establish a base rate for bar pressing behavior. The *S*s were then matched by bar pressing rate, within each strain separately, and randomly assigned to the two groups of seven *S*s each. This procedure resulted in a close matching (bar pressing *M*s of 58.1 and 58.7 for the 8 trials).

Testing.—Group I received increasing concentrations (4, 8, 32, and 64%) of sucrose in water during four spaced training trials; Group D received decreasing concentrations (64, 32, 8, and 4%).

The daily procedure, repeated over 12 successive days, consisted of four 30-sec. training trials followed by four 30-sec. nonreinforced trials, with 120-sec. intertrial intervals (light on, bar locked). All bar presses were reinforced. During the nonreinforced trials, the *S*s bar pressed for a dry dipper. They were not removed from the boxes at any time during the sequence of eight trials. The four albino *S*s from Group I and the three hooded *S*s from Group D were tested first, followed by the three hoods from Group I and the four albinos from Group D.

Results

Figure 1 shows the group means for each of the four reinforced and four nonreinforced trials averaged over the 12 test days. An examination of this curve indicates that the results were as predicted. Group I had a lower terminal reinforced response rate than Group D, but this relationship was reversed in the following nonreinforced trials. Very similar relationships were evident throughout the course of the 12-day test period when the same points were plotted in blocks of 3 days, indicating that the effect was not peculiar to any phase of the experiment.

Total bar presses, summed for the 12 days, were subjected to an analysis of variance appropriate for a three-factor experiment in which there are repeated measures on two factors. The variables tested were those shown in Fig. 1: order (experimental treatment), reinforcement, and trials. All main effect *F*s were highly reliable. The interaction between reinforcement and order is the most important comparison in terms of the major prediction of the experiment. This *F* was highly reliable: $F(1, 12) = 33.48$, $p < .005$.

A further breakdown of the analysis in which the appropriate error terms from the major analysis were utilized showed that there was a reliable difference in group means for the reinforced trials: $F(1, 12) = 39.76$, $p < .005$, with Group D having a higher level of performance. Also, the superior performance of Group I on nonreinforcement trials was reliable: $F(1, 12) = 6.15$, $p < .05$.

The interaction between the last reinforced trial and the first nonreinforced trial was computed and found to be reliable: $F(1, 12) = 25.38$, $p < .005$, indicating that there was a substantial shift between terminal reinforced performance and initial nonreinforced behavior.

No interaction between order and trials occurred over the test days. There was simply a linear decrease in nonreinforced responding for both groups, with Group I showing a consistently higher rate.

Individual means for reinforced and

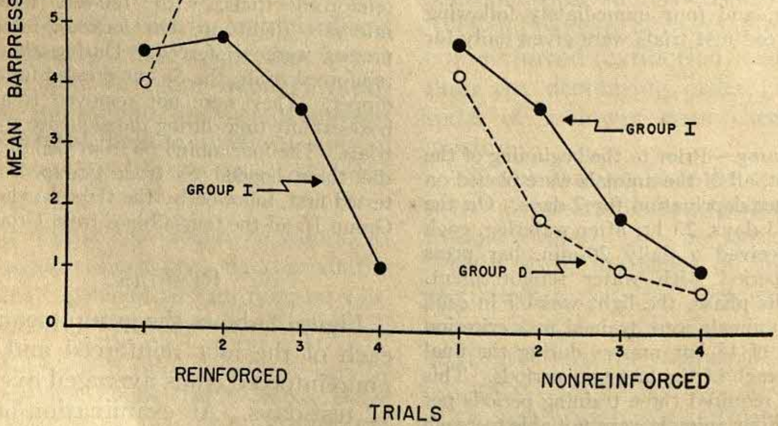


FIG. 1. Mean bar presses on reinforced trials and nonreinforced trials.

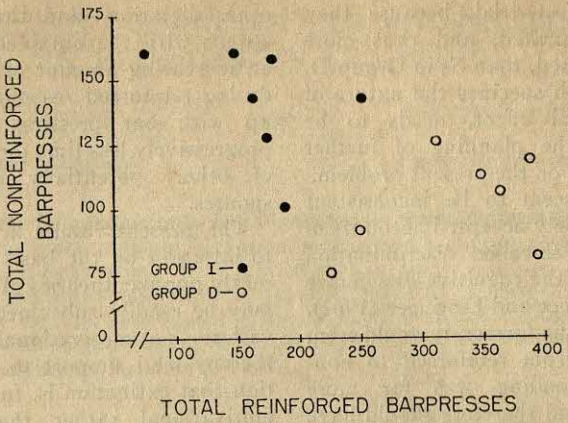


FIG. 2. Scattergram of individual bar pressing totals on reinforced and nonreinforced trials.

nonreinforced bar presses are plotted in Fig. 2. The clear separation of the two groups indicates that the differences evaluated above in terms of group means were generally true of the individual *S*s.

DISCUSSION

The major theoretical significance of the present results is the demonstration of the necessity for assuming some kind of a nonassociative—that is, motivational—factor to account for the differential persistence of nonreinforced responding. More specifically, these data give support to the interpretation of extinction as a reduction in performance that is essentially due to a motivational decrement (Marx, 1958, 1960) rather than merely the terminal level of reinforced performance. The present test is a particularly rigorous one of the motivational account because of the great weight typically accorded rate of reinforced responding as a determiner of nonreinforced response rate. In a large number of experimental investigations of extinction (Pubols, 1960) this has proved to be by far the most significant variable affecting resistance to extinction.

These conclusions depend of course upon the assumption that Group I *S*s *were* more strongly motivated at the end of the last reinforced trial, bar pressing for a highly preferred (64%) sucrose concentration, than Group D *S*s. Although this assumption certainly seems to be a reasonable one, further research should have some direct and independent measure of motivation.

Whether the increasing order is necessary for the observed effect, or the differential terminal concentrations (64% compared with 4%) are sufficient, will also need to be determined in further research. In either case a motivational account would seem to be necessary, as opposed to an explanation on the basis of terminal level of reinforced performance.

The results of this experiment may also be explained on the basis of one variety of frustration theory, which offers a third motivational account. This account requires two assumptions: first, that it is more frustrating to shift from a high, desirable incentive (64%) than from a weak incentive (4%) to no incentive, which seems entirely reasonable; and second, that there is a substantial positive correlation between such frustration and the vigor of associated responses, which is popularly accepted but is not unequivocally supported by controlled data (Lawson & Marx, 1958; Marx, 1956). In other words, Group I *S*s responded more vigorously, especially during the first nonreinforced test trial, because they were more frustrated, and thus more strongly motivated, than *S*s in Group D. This view, which specifies the nature of the motivational effect, needs to be considered in the planning of further experimentation on the present problem.

Our results seem to be inconsistent with certain widely accepted accounts of extinction: the so-called discrimination hypothesis, and the cognitive-dissonance theory of Lawrence and Festinger (1962). With regard to the former, it would seem that the shift from reinforced to nonreinforced responding was far more discriminable, and therefore should have resulted in faster extinction, for Group I (64–0%) than for Group D (4–0%). The fact that exactly the opposite results occurred is difficult to reconcile on the basis of a discrimination account of extinction.

The cognitive-dissonance notion accounts for more extended nonreinforced performance on the basis of a greater discrepancy between expectation and incentive outcome. The present experiment, in which one group found decreasingly desirable and the other group increasingly desirable incentives, would seem to provide something of a test for this view. However, the results again were exactly opposite to those which would follow from cognitive-dissonance theory: Group D *S*s, with widening discrepancy between expectation and incentive, should have had the greater dissonance and therefore the higher nonreinforced response rate.

The competing-response interpretation of extinction does not seem to be directly applicable to the present experiment. This view, itself in need of further specification (cf. Marx, in press), requires the presence of responses incompatible with the bar press response. It is not at all apparent why such responses, if indeed they did occur in the present experiment, should have been more prominent in Group D than Group I during nonreinforced tests. The difference, if any, should have been in the other direction—and thus again inconsistent with the obtained results—since an increasing amount of Group D time during reinforced responding was taken up with bar pressing, thus allowing progressively less time for the acquisition of other, potentially competing, responses.

The present results are thus difficult to interpret on the basis of several currently popular theories of extinction, but may be readily subsumed under several variations of motivational theory. They thereby lend support to the generalization that extinction is, fundamentally, a motivational rather than strictly an associative process.

REFERENCES

Collier, G., & Willis, F. N. Deprivation and reinforcement. *J. exp. Psychol.*, 1961, **62**, 377–384.

Lawrence, D. H., & Festinger, L. *Deterrents and reinforcement.* Stanford: Stanford Univer. Press, 1962.

Lawson, R., & Marx, M. H. Frustration: Theory and experiment. *Genet. Psychol. Monogr.*, 1958, **57**, 393–464.

Marx, M. H. Some relations between frustration and drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1956.* Lincoln: Univer. Nebraska Press, 1956.

Marx, M. H. Resistance to extinction as a function of continuous or intermittent presentation of a training cue. *J. exp. Psychol.*, 1958, **56**, 251–255.

Marx, M. H. Resistance to extinction as a function of degree of reproduction of training conditions. *J. exp. Psychol.*, 1960, **59**, 337–342.

Marx, M. H. The need for more specific formulations of the "competing-response" interpretation of extinction. *Psychol. Rep.*, in press.

Marx, M. H., & Murphy, W. W. Resistance to extinction as a function of the presentation of a motivation cue in the startbox. *J. comp. physiol. Psychol.*, 1961, **54**, 207–210.

Pubols, B. H. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

(Received December 10, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 547–552

SEQUENTIAL DEPENDENCIES AND NONREINFORCEMENT IN PROBABILITY LEARNING[1]

JAMES G. GREENO AND DAVID LaBERGE
Indiana University *University of Minnesota*

To obtain information regarding dependence of nonreinforcement (E_0) effects on other variables, E_0's were presented with varying response-contingent frequencies. With 24 student Ss in each of 8 groups, mean choice frequencies and sequential statistics indicated (a) that S tends to perseverate on the response chosen on an E_0 trial, and (b) that the effect of an E_0 preceded by E_1 in the reinforcement series is different from the effect of E_0 preceded by E_2.

When *S* is given the choice of predicting which of two events will occur and neither of the events occurs, what effect does this have on his future choice behavior? This question has puzzled numerous investigators of probability learning (Anderson & Grant, 1957, 1958; Atkinson, 1956; Greeno, 1962; LaBerge, Greeno, & Peterson, 1962; Neimark, 1956) and a variety of models have been proposed and tested. However, about all that we can confidently say about the effects of these nonreinforced trials is that they sometimes cause *S*'s choice frequencies to remain closer to 50:50 than would be the case if nonreinforcement were not present (Estes, in press).

The present experiment was designed to provide information regarding certain sequential dependencies in the effects of nonreinforced trials in probability learning. We believe that general information is needed regarding the extent to which the effects of nonreinforced trials depend on other variables in the situation before we can hope to describe the effects of nonreinforcement in adequate detail.

METHOD

Design.—Groups were run with probability learning schedules in which the relative frequencies of the two reinforcement events, E_1 and E_2, were 50:50 or 75:25. These relative frequencies were noncontingent, in the sense that the ratio of E_1, to E_2 frequencies was the same for trials on which *S* predicted E_1 as for trials on which *S* predicted E_2. However, in some groups, the frequency with which nonreinforced or E_0 trials occurred was different depending on *S*'s choices.

Descriptions of the sequences received by *S*s in the various experimental groups are presented in Table 1. The number of trials per block was varied so that each group would receive approximately 20 reinforced trials per block (cf. Greeno, 1962; LaBerge et al., 1962). Note that groups labeled "50/" received E_1:E_2 ratios of 50:50; groups labeled "75/" received E_1:E_2 ratios of 75:25. The notation "/0" refers to groups who received no E_0 trials; "/N" denotes groups for which E_0 was equally likely to follow A_1 choices (predictions of E_1) as A_2 choices (predictions of E_2); "/2" denotes groups for which E_0 was more likely following A_2 than A_1; and "/1" denotes groups for which E_0 was more likely to follow A_1 than A_2. The frequencies of E_0 were selected on the basis of previous results (Greeno, 1962) in order to provide approximately equal overall frequencies of E_0, or $\bar{\pi}_0$, for groups which were to be compared. This consideration led to

---

[1] This research was supported in part by a grant from the National Science Foundation to the second author. It was also supported in part by Grant No. G-5525 from the National Science Foundation and by Contract 908(16) between the Office of Naval Research and Indiana University, for which W. K. Estes was Principal Investigator. We gratefully acknowledge the helpful suggestions provided by Norman H. Anderson regarding this report.

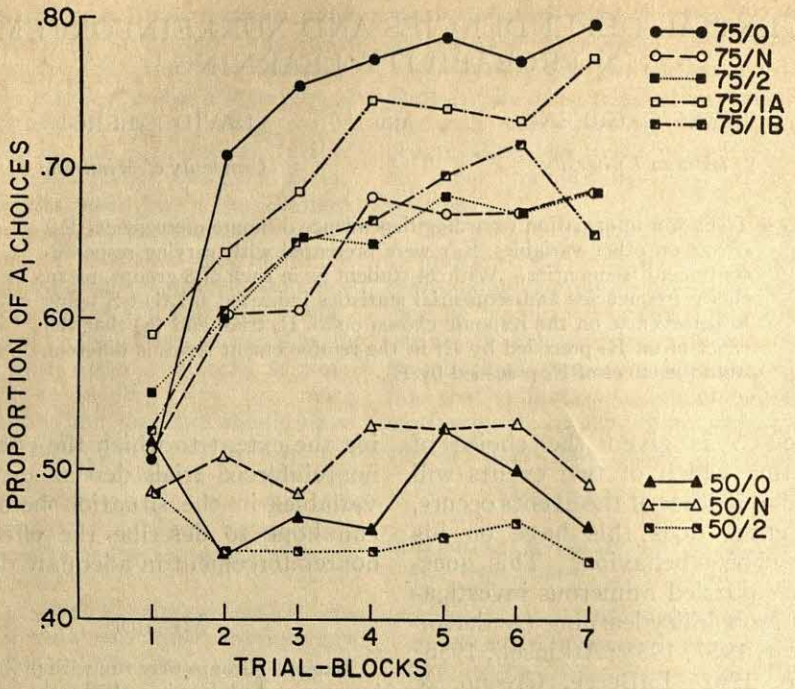


FIG. 1. Proportions of A_1 choices for groups varying in E_1:E_2 ratio and in contingent frequencies of E_0.

the use of higher frequencies of E_0 in Blocks 1–2 for Group 75/1B, because of the fact that the frequencies used for Group 75/1A were certain to produce a low value of $\bar{\pi}_0$ in early trials.

Procedure.—The 192 *S*s were students in summer session psychology courses at the University of Minnesota. The 50:50 groups were run with 24 *S*s randomly allocated to each of the groups, after which 24 *S*s were randomly allocated to each of the 75:25 groups.

Two *S*s at a time were seated at a divided table where slanting panels with two buttons and two lights were fastened. The *S*s were instructed to predict which of the two lights would flash on each trial by pressing one of the buttons. Eight practice trials were given: $E_1E_1E_2E_2E_0E_0E_1E_2$. If *S* remarked or asked about the E_0 trials, he was told, "That may happen on some trials. In any case, you must make a choice on each trial." No other instructions were given regarding the E_0 trials. The sequence of events on each trial was: buzzer signal, 2 sec.; off, 1 sec.; reinforcement event, 1 sec.; off, 2 sec. The temporal sequence of events was automatically controlled. The *S*s' choices and the reinforcement event for each trial were automatically recorded.

Sequences of reinforcement events were constructed by obtaining random orderings within blocks of trials each of which included about 20 reinforced trials. Schedules were matched across groups in so far as was possible. The side designated A_1 was counterbalanced. Each group received seven blocks of trials, and trials were presented without interruption.

RESULTS

Mean proportions of A_1 choices, or P_1, are graphed by blocks in Fig. 1. Values of P_1 were obtained for individual *S*s over Blocks 4–7, and means of these scores are presented in Table 1. Estimates of σ^2 of .0096 ($df = 36$) and .0135 ($df = 60$) were obtained for the 50:50 and 75:25 groups, respectively, as residual mean square terms in factorial analyses of variance with experimental groups, reinforcement schedules, and A_1 sides as the factors. These estimates were used to compute the 90% confidence intervals of orthogonal contrasts to be described below.

The overall effect of adding E_0 trials to the sequences was estimated by comparing P_1 from a /0 group with the mean P_1 from the groups with the same E_1:E_2 ratio. For the 75:25 groups, the estimated effect on P_1 due to presence of E_0 trials was $-.085 \pm 0.43$, which is about the same as that reported earlier (Greeno, 1962) for groups with differing incentives for the two responses. For the 50:50 groups, the resulting estimate was $-.001 \pm .041$, indicating that the E_0 trials produced little or no overall change. The effects of varying the frequency of E_0 depending on *S*'s choices in the 75:25 groups were estimated by comparing P_1 of Groups 75/2 and 75/1B. The resulting estimate was $-.011 \pm .056$, indicating that there was little or no effect on P_1 due to varying the choice which was more likely to be followed by E_0. For the 50:50 groups, the appropriate estimate is the difference between Groups 75/2 and 75/N, which was $-.069 \pm .047$. That is, with E_1 and E_2 occurring equally often, *S*s showed a preference for the choice which was more frequently followed by E_0.

Further information regarding sequential dependencies can be obtained from the sequential statistics in Table 2. The relative frequencies of joint events $(E_{h,n-1}A_{i,n}E_{j,n}A_{k,n+1})$ in Blocks 4–7 have been divided by the conditional frequencies: $P(E_{j,n}A_{i,n})$ in order that we may use the fact that:

$$\frac{P(A_{2,n}E_{j,n}A_{1,n+1})}{P(E_{j,n}|A_{2,n})} - \frac{P(A_{1,n}E_{j,n}A_{2,n+1})}{P(E_{j,n}|A_{1,n})} = P(A_{1,n+1}|E_{j,n}) - P(A_{1,n}) + X_{j,n} \qquad [1]$$

The $X_{j,n}$ term is complex; however, it equals zero whenever $P(E_{j,n}|A_{1,n}) = P(E_{j,n}|A_{2,n})$, which is true for Groups /0 and /N. When $X_{j,n}=0$, the expression in Equation 1, or $D(E_j)$, provides an unbiased estimate of the

TABLE 1

DESCRIPTION OF REINFORCEMENT SEQUENCES AND OBTAINED VALUES OF P_1

Group	Trials per Block		E_1	E_2	E_0	$\bar{\pi}_0$		P_1 Blocks 4–7
						Blocks 1–2	Blocks 4–7	
50/0	20		10	10	0	.000	.000	.486
50/N	33		10	10	13	.394	.394	.519
50/2	33	if A_1	15	15	3	.415	.424	.451
		if A_2	5	5	23			
75/0	20		15	5	0	.000	.000	.780
75/N	35		15	5	15	.429	.429	.677
75/2	35	if A_1	21	7	7	.446	.384	.672
		if A_2	9	3	23			
75/1B	35 Blocks 1–2	if A_1	9	3	23	.458	.398	.683
		if A_2	21	7	7			
	Blocks 3–7	if A_1	12	4	19			
		if A_2	24	8	3			
75/1A	35	if A_1	12	4	19	.345	.427	.747
		if A_2	24	8	3			

TABLE 2

FREQUENCIES OF 4-TUPLE $(E_hA_iE_jA_k)$ EVENTS DIVIDED BY π_{ij}

4-Gram			Group							
$n-1$	n	$n+1$	50/0	50/N	50/2	75/0	75/N	75/2-	75/1B	75/1A
1	10	1		.111	.099		.238	.239	.224	.243
1	10	2		.079	.088		.087	.122	.103	.099
1	20	1		.055	.046		.039	.054	.074	.049
1	20	2		.061	.071		.050	.039	.060	.063
2	10	1		.063	.085		.049	.053	.053	.058
2	10	2		.055	.067		.015	.026	.020	.018
2	20	1		.086	.062		.027	.034	.035	.060
2	20	2		.081	.080		.038	.026	.056	.025
0	10	1		.131	.085		.199	.146	.205	.253
0	10	2		.083	.063		.077	.056	.075	.059
0	20	1		.083	.072		.076	.069	.049	.035
0	20	2		.100	.197		.090	.103	.063	.049
d_1	(E_0)			*−.024*	*−.042*		*−.048*	*−.068*	*−.029*	*−.050*
d_2	(E_0)			*+.031*	*−.005*		*+.012*	*+.008*	*+.015*	*+.042*
d_0	(E_0)			*.000*	*+.009*		*−.001*	*+.013*	*−.026*	*−.024*
D	(E_0)			**+.007**	**−.038**		**−.037**	**−.047**	**−.040**	**−.032**
1	11	1	.165	.119	.098	.500	.289	.312	.251	.325
1	11	2	.121	.034	.066	.089	.051	.057	.066	.036
1	21	1	.089	.052	.057	.100	.050	.049	.067	.051
1	21	2	.116	.073	.076	.056	.040	.057	.051	.029
2	11	1	.128	.069	.068	.147	.049	.079	.057	.071
2	11	2	.101	.056	.061	.045	.022	.023	.018	.015
2	21	1	.180	.108	.082	.058	.057	.039	.051	.046
2	21	2	.120	.066	.068	.015	.020	.025	.021	.013
0	11	1		.128	.097		.224	.168	.218	.247
0	11	2		.102	.059		.046	.040	.048	.056
0	21	1		.121	.157		.107	.114	.087	.087
0	21	2		.083	.129		.044	.040	.039	.026
d_1	(E_1)		*−.032*	*+.018*	*−.009*	*+.011*	*−.001*	*−.008*	*+.001*	*+.015*
d_2	(E_1)		*+.079*	*+.052*	*+.021*	*+.013*	*+.035*	*+.016*	*+.033*	*+.031*
d_0	(E_1)			*+.019*	*+.098*		*+.061*	*+.074*	*+.039*	*+.031*
D	(E_1)		**+.047**	**+.089**	**+.110**	**+.024**	**+.095**	**+.082**	**+.073**	**+.077**
1	12	1	.096	.086	.088	.490	.189	.239	.166	.206
1	12	2	.171	.113	.068	.155	.167	.140	.143	.129
1	22	1	.107	.058	.066	.103	.054	.037	.067	.049
1	22	2	.137	.078	.082	.028	.069	.026	.057	.035
2	12	1	.115	.077	.070	.116	.026	.059	.050	.047
2	12	2	.075	.061	.061	.037	.039	.032	.031	.034
2	22	1	.102	.045	.053	.011	.024	.042	.021	.014
2	22	2	.178	.130	.089	.032	.052	.048	.036	.040
0	12	1		.086	.066		.135	.130	.158	.217
0	12	2		.087	.083		.103	.077	.103	.108
0	22	1		.064	.095		.052	.063	.048	.042
0	22	2		.098	.125		.064	.106	.066	.049
d_1	(E_2)		*−.064*	*−.055*	*+.002*	*−.052*	*−.113*	*−.103*	*−.076*	*−.080*
d_2	(E_2)		*+.027*	*−.016*	*−.008*	*−.026*	*−.015*	*+.010*	*−.010*	*−.020*
d_0	(E_2)			*−.023*	*+.012*		*−.051*	*−.014*	*−.055*	*−.066*
D	(E_2)		**−.037**	**−.094**	**+.002**	**−.078**	**−.179**	**−.109**	**−.141**	**−.116**

average change in P_1 across E_j trials (Anderson & Grant, 1958). $X_{j,n}$ will usually be positive when $P(E_{j,n}|A_{1,n}) < P(E_{j,n}|A_{2,n})$ as in the case for E_0 in Groups /2. $X_{j,n}$ will usually be negative when $P(E_{j,n}A_{1,n}) > P(E_{j,n}A_{2,n})$ as is the case for Groups 75/1-.

The 4-tuple frequencies were obtained in order to see whether the average change in P_1 across E_0 trials differed depending on whether E_0 had been preceded by E_1, E_2, or E_0. The value of $D(E_j)$ is, of course, an average over all trials and in examining the 4-tuple statistics, it was found that the statistics obtained by partitioning trials on the basis of $E_{h,n-1}$ were markedly different. These separate statistics are presented in Table 2, denoted $d_h(E_j)$, in italics, with $D(E_j)$'s in bold type. The $d_h(E_j)$'s can be taken as estimates of the average change in P_1 across E_j trials which followed E_h trials and are subject to the same biases as those described above for $D(E_j)$.

Because of the above-mentioned biases in $D(E_j)$ for the contingent groups, we focus primarily on the noncontingent /N and /0 groups. The data in Table 2 reveal two important facts. First, P_1 decreased across E_0 trials preceded by E_1; but P_1 increased across E_0 trials preceded by E_2. Secondly, P_1 increased substantially across E_1 trials which followed E_2 or E_0; but P_1 either increased by smaller amounts or decreased across E_1 trials preceded by E_1. The latter finding is consistent with earlier reports of negative recency (Anderson, 1960; Anderson & Whalen, 1960; Jarvik, 1951; Nicks, 1959). The findings regarding changes across E_0 trials suggest another form of negative recency—one in which E_0 trials produce a decrease in the probability of the more recently reinforced response.

The general pattern of the sequential statistics for the /2 and /1- groups is much the same as that for the noncontingent groups. However, the biases in these statistics must be taken into account in interpreting them, and this is best deferred to the next section.

DISCUSSION

Sequential dependencies in the effects of E_0 trials could be based on either or both of two variables—S's choices, or the events in the series of reinforcements. The design of the experiment was intended to make it especially sensitive to dependencies involving S's choices, in that the frequencies with which E_0 followed the two choices were varied among groups. The absence of between-groups differences in P_1 due to this variable among the 75/ groups, and the similar pattern of sequential statistics among these groups would suggest that if E_0 effects depend on S's choices, then this dependence is probably weak.

However, the value of P_1 for Group 50/2 and the biases in $D(E_0)$ for the contingent groups suggests a response-dependent sequential effect. The 50/2 result suggests that S may perseverate his choice on Trial n when he receives E_0. When the probable biases described above for $D(E_j)$ are taken into account, the data suggest that P_1 decreased more across E_0 trials in Group 75/2 than in 75/1B, and this would be consistent with a response-perseveration effect. Response perseveration should cause P_1 to be higher in the 75/1B schedule than in the 75/2 schedule, and this was not observed; but it is possible that such a difference might have occurred had the sequences been continued beyond Block 7.

The finding that the effect of E_0 depended on recent reinforcement events is more conclusive. The difference between $d_1(E_0)$ and $d_2(E_0)$ was consistent across all groups; and, although the results from the contingent groups are not conclusive because of the possibility of differential biases, the presence of the same effect in these groups adds weight to the conclusion that E_0 following E_1 has an effect different from that of E_0 following E_2. Further work is needed to specify the nature of response perseveration and sequential event dependencies in E_0 effects; but the presence of these processes in the results of this experiment may serve as useful guidelines to such investigations.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317.

ANDERSON, N. H., & GRANT, D. A. Correction and reanalysis. *J. exp. Psychol.*, 1958, **56**, 453–454.

ANDERSON, N. H., & WHALEN, R. E. Likelihood judgments and sequential effects in a two-choice probability learning situation. *J. exp. Psychol.*, 1960, **60**, 111–120.

ATKINSON, R. D. An analysis of the effect of nonreinforced trials in terms of statistical learning theory. *J. exp. Psychol.*, 1956, **52**, 28–32.

ESTES, W. K. Probability learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, in press. Ch. 4.

GREENO, J. G. Effects of nonreinforced trials in two-choice learning with noncontingent reinforcement. *J. exp. Psychol.*, 1962, **64**, 373–379.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

LABERGE, D., GREENO, J. G., & PETERSON, O. F. Nonreinforcement and neutralization of stimuli. *J. exp. Psychol.*, 1962, **63**, 207–213.

NEIMARK, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

(Received December 19, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 553–559

EFFECTS OF ALTERED FREQUENCIES UPON RECOGNITION RESPONSES[1]

ARNOLD BINDER

Indiana University

The experiment investigated 2 aspects of the effects of a shift in the relative frequencies of stimuli during paired-associate learning upon response frequencies to cues common to these stimuli. It was found that the frequencies of the initial 2 or 4 blocks (the shift occurred after 2 blocks for some groups and after 4 for others) affected test responses even after 10 or 12 blocks at the altered frequencies. Direct support was not found for the earlier reported results that superior learners tend to match the early, rather than the later, frequencies in their responses to common elements, although there was a tendency apparent for superior learners to be more influenced by the initial frequencies than inferior learners.

Binder and Feldman (1960) investigated the phenomenon of differential cue selection (or selective sampling) in an experiment involving the learning of the names of figures containing both common and discriminative cues. The relative frequencies of occurrence of these figures were markedly shifted at a fixed point during a sequence of paired-associate trials. Following learning to a stringent criterion, *S*s were given test trials during each of which they responded to ambiguous cue-components of the figures with one of the previously learned names. Under the assumption that superior learning ability is associated with the sampling of discriminative cues only, support was obtained for the hypothesis that the pattern of response frequencies to ambiguous cues is a function of differential attention among the common and discriminative components during prior acquisition. In particular, *S*s who made no errors after reaching a preliminary criterion responded to common, but not discriminative, cues more nearly in accord with the initial learning frequencies, while *S*s who made errors tended to match the later or altered frequencies. Similar results were reported by LaBerge and Smith (1957).

The present experiment was aimed at a more systematic and extensive investigation of this differential selection hypothesis as well as the investigation of the overall effectiveness in test response determination of early vs. later learning frequencies. Some evidence that the effects of an early reinforcement schedule linger for a long while after the schedule is altered has been provided by Anderson (1960).

METHOD

Subjects.—The *S*s, all undergraduates at Indiana University, were run in five separate groups. The numbers of *S*s were as follows: Group I, 113; Group II, 59; Group III, 75; Group IV, 59; Group V, 47.

Apparatus.—The stimulus materials were projected onto a matte screen by two random access projectors. (The programing of the presentation was semiautomatic; *E* fixed dials which controlled the sequence of stimuli for two blocks in advance.) Four booths were arranged in an arc so that *S*s sitting in them had approximately equally adequate visibility of the screen. The booths were open in the front, for entering, and in the rear, for viewing the screen, and were lined with acoustical tile on the sides and top. There was a desk in each booth with a hole in the center. A roll of Esterline-Angus paper passed directly under the hole and *S* wrote his response on this paper through the hole. Movement of the paper was in synchrony with stimulus presentation and removal.

---

[1] This research was supported by Research Grant M-2170 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

Materials.—Four nonsense syllables (WOY, KAL, RUX, and VED) of 33% association value (Glaze, 1928) were used as the names during the learning trials. Four stimulus figures were also used and may be seen in Fig. 1. In addition to these learning figures, Fig. 1 contains the test stimuli. Special care was taken to assure that identical cues on the different stimuli appeared in the same location on the projection screen.

Procedure.—All *S*s learned the figure-name associations by anticipation (with correction), with figures appearing to their left and names to their right on the screen. The following intervals were used: stimulus exposure, 5½ sec.; response exposure, 2 sec.; intertrial interval, 7 sec.; and delay after warning buzzer, 1 sec. The interblock interval was 14 sec.

The learning blocks for *S*s of Group I were eight paired associates in length. One of the C_i figures (Fig. 1) occurred twice, one four times, and the other two once each. It was randomly determined which of the two pairs C_1-C_2 and C_3-C_4 would occur at the 2:1 ratio and which at the 4:1 ratio, and then a second random determination was made as to which figure, in each pair, occurred at the higher frequency and which at the lower frequency. Within each odd-numbered trial block the order of the appearance of the pairs was randomized with the restriction of no more than two consecutive occurrences of a given stimulus. The reverse order of presentation to that of the immediately preceding block was used for each even-numbered block. Twenty-four such randomizations of frequency and order were prepared.

There were eight learning blocks, a rest period of 3 min., and two more learning blocks. A test trial consisting of the presentation of C (Fig. 1) followed these last two blocks. Prior to the test presentation *S*s were instructed to give "the name that seems most appropriate." The test stimulus was shown for 2 sec., beginning 1 sec. after the sounding of a buzzer.

Sign-ups for the experiment were in subgroups of four (24 subgroups in all), but for the usual array of reasons it was necessary to run groups with fewer than four *S*s present. Another important source of attrition was the breakdown of equipment which at times did not become evident until the data were being read. The total number of *S*s was so depleted by the attrition that it was decided to rerun those randomizations (or subgroups) that contained fewer than four *S*s. It was further decided that it would be most efficient to get *S*s for this process of filling in the gaps in subgroups of four because of the likelihood of further attrition. Consequently, in the final sample some subgroups have more than

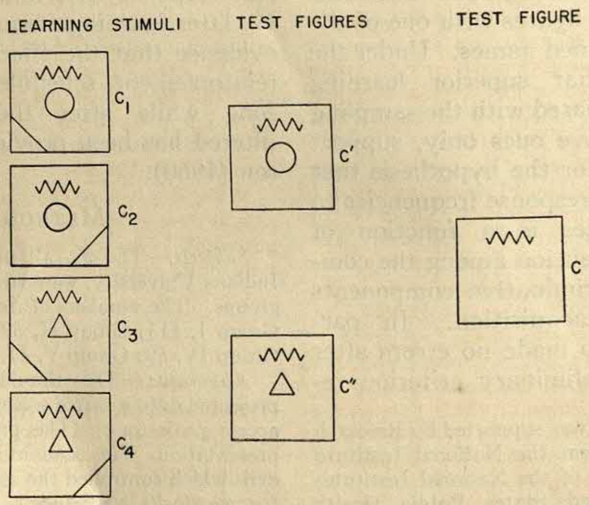


FIG. 1. The stimuli used for learning and testing.

TABLE 1

Composition of Acquisition Blocks

Randomized Sequence of Blocks[a]	Stimulus Frequency per Block				Block Ordering for Groups of *S*s				
	C_i	$C_{i'}$	C_j	$C_{j'}$	I	II	III	IV	V
A	1	1	1	1			1	1	
B	1	1	1	1			2	2	
C	1	1	1	1		1	3	3	1
D	1	1	1	1		2	4	4	2
E	2	1	4	1	1	3	5	5	3
F	2	1	4	1	2	4	6	6	4
G	2	1	4	1	3	5	7	7	5
H	2	1	4	1	4	6	8	8	6
I	2	1	4	1	5	7	9	9	7
J	2	1	4	1	6	8	10	10	8
K	2	1	4	1	7	9	11	11	9
L	2	1	4	1	8	10	12	12	10
3-Min. Rest Period									
M	2	1	4	1	9	11	13	13	11
N	2	1	4	1	10	12	14	14	12
Test Stimulus Exposure									
O	2	1	4	1				15	13
P	2	1	4	1				16	14
Test Stimulus Exposure									

[a] There was a total of 24 such sequences randomized for both frequency of particular stimuli and order of presentation per block.

four members (with a maximum of seven) and the total number exceeds 96 (24 × 4).

During the first two learning blocks of Group II, each of the learning figures occurred once. Then, beginning with the third block the 24 subgroups in this experiment underwent training identical to that of Group I. The same randomizations and frequencies were followed for each subgroup in Group II as the comparable subgroup in Group I. Thus, the total learning experience for *S*s in Group II consisted of two blocks of equal frequency, eight blocks of 2:1:4:1, rest period, two blocks of 2:1:4:1.

This manner of equating blocks is illustrated in Table 1. The 10 blocks for Group I are shown in the extreme left column of the third (and last) grouping of columns. The frequencies of the stimuli are 2:1:4:1 over all blocks for this group, as is indicated in the second grouping of columns. Block 3 of Group II is identical to Block 1 of Group I, Block 4 of II to Block 2 of Group I, etc., and so they are shown on the same rows of the table. The first column of Table 1 shows the complete randomized order for the group with the largest number of acquisition trials, to be discussed below.

Group III was like Group II except that four, rather than two, blocks of equal frequencies preceded the 2:1:4:1 shift. The last two of the four equally frequent blocks for the 24 subgroups in Group III were identical to the first two blocks of the comparable subgroups in Group II (see Table 1).

The *S*s in both Groups II and III were tested with Stimulus C. Group V was in every way identical to Group II, and Group IV to Group III, up to the presentation of the test stimulus (see Table 1). In Groups IV and V, either Test Stimulus C′ or C″ (as randomly determined) was given in place of C; two more blocks at 2:1:4:1 followed, and then another test consisting of the C′ or C″ not shown on the first test.

While it was hoped to remove the attrition gaps in these groups as in Group I, this proved impossible because of time, apparatus, and *S* availability limitations.

Results

The learning curves in terms of errors per attempted anticipation may be seen in Fig. 2. The numbers for the learning blocks are not necessarily the same as the sequence numbers of the blocks for a particular group. The

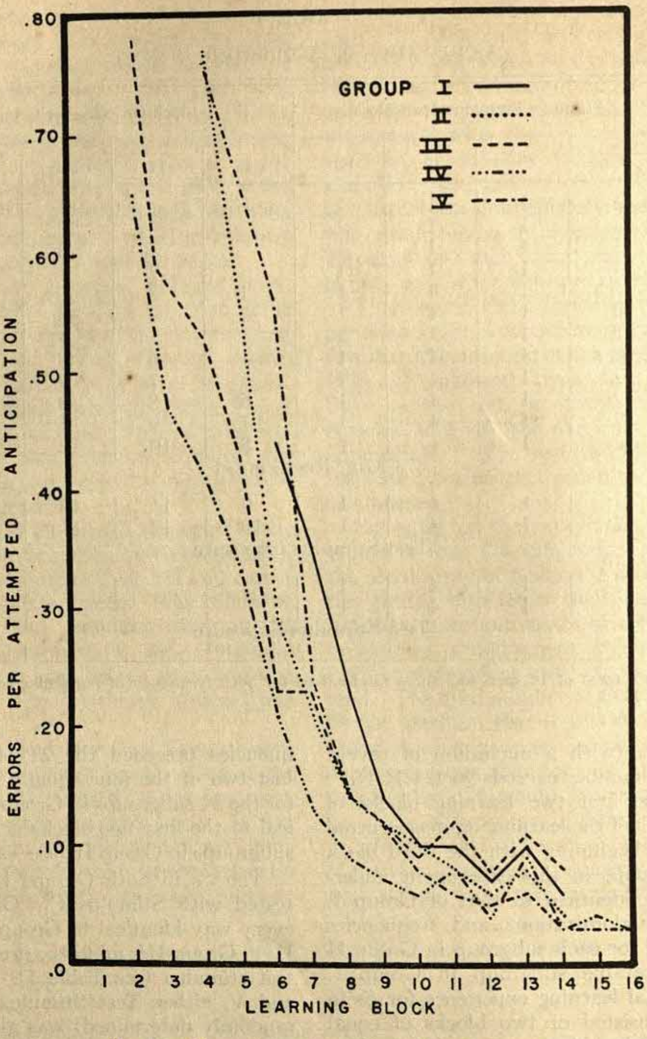


FIG. 2. Learning curves for the five groups.

numbering of blocks in Fig. 2 actually reflects the ordering of blocks for Group IV, but the plotting for the other groups is in terms of the blocks corresponding (in frequency and randomization) to the blocks of Group IV as shown in Table 1. Thus, for example, Groups II and IV differ both in number of blocks of equal frequency (2 for II and 4 for IV) and in the number of blocks following the rest period (2 for II and 4 for IV). Therefore, the plotting for Group II in Fig. 2 starts with the block numbered 4 (the first block for each group was not plotted) and ends with the block numbered 14. The block numbered 14 in Fig. 2 is actually the twelfth block of Group II.

In accord with the usual procedure in this type of investigation it was necessary to pick out those individuals who were making the discriminations nearly perfectly prior to their test trials. Failure to sample selectively could of course lead to an error. Accordingly, *S*s who achieved at least two blocks of perfect anticipations prior to a given test trial were considered superior learners for this purpose. This seemed to be a reasonably high criterion and yet gave approximately equal numbers of *S*s in the superior and inferior learning ability categories.

The responses to the test stimuli of superior and inferior learners in the five groups may be seen in Table 2. The test responses consisting of the names of learning figures which appeared twice and four times have been pooled into a high frequency category and the two figures which appeared once each into a low frequency category. This gave larger frequencies without sacrificing important information since the critical factor in the analysis is the difference between equal and differential frequency responding.

Below the observed frequencies, within each group, are the frequencies predicted from each of a number of matching bases. Matching on the basis of the equally frequent early blocks leads to the expectation of an equal number of high and low frequency responses while matching on the basis of the differential frequency blocks leads to the expectation of about a 3 to 1 ratio of high to low frequency test responses. (Exactly 3:1 for Groups II and III and about 2.75 to 1, depending upon the respective numbers of responses to C′ and C″, for Groups IV and V.) The row labeled "Overall matching" refers to the frequencies expected if *S*s matched the frequencies as summed over all learning blocks for each figure. For example in Group III the overall relative frequencies for the four figures

TABLE 2

OBTAINED AND EXPECTED RESPONSE FREQUENCIES

Group	Blocks Prior to Shift	Entry Basis	Inferior Learners		Superior Learners	
			High	Low	High	Low
I	None	Observed	42	14	45	10
		Matching 4:1:2:1	42.0	14.0	41.2	13.8
II	2	Observed	25	7	16	11
		Matching 1:1:1:1	16.0	16.0	13.5	13.5
		Matching 2:1:4:1	24.0	8.0	20.2	6.8
		Overall matching	23.3	8.7	19.6	7.4
III	4	Observed	28	11	24	12
		Matching 1:1:1:1	19.5	19.5	18.0	18.0
		Matching 2:1:4:1	29.2	9.8	27.0	9.0
		Overall matching	27.6	11.4	25.5	10.5
IV	4	Observed	22	13	40	26
		Matching 1:1:1:1	17.5	17.5	33.0	33.0
		Matching 2:1:4:1	25.7	9.3	48.5	17.5
		Overall matching	24.7	10.3	46.1	19.9
V	2	Observed	22	11	35	19
		Matching 1:1:1:1	16.5	16.5	27.0	27.0
		Matching 2:1:4:1	24.1	8.9	39.5	14.5
		Overall matching	23.4	9.6	38.6	15.4

were $\frac{44}{96}$, $\frac{24}{96}$, $\frac{14}{96}$, and $\frac{14}{96}$. This compares with proportions of $\frac{40}{80}$, $\frac{20}{80}$, $\frac{10}{80}$, and $\frac{10}{80}$ when expectations are computed from the latter (2:1:4:1) blocks alone.

Errors during testing, such as omissions and the name of C_3 or C_4 when C' is shown, have been excluded from the tabulation. The total numbers of these errors for the various groups are as follows: Group I, 2; Group II, 0; Group III, 0; Group IV, 17; Group V, 7.

Comparing the observed frequencies against "Matching 2:1:4:1" and "Overall matching," we see that in only one such comparison (out of eight) is the fit closer to the 2:1:4:1 expectations than to the overall expectations. The probability of this event (or a more extreme one) is less than .0312 under the composite hypothesis that, in any given comparison, "Matching 2:1:4:1" is at least as likely as "Overall matching" to provide a better fit, with the size bounded above.

According to the selective sampling hypothesis, perfect discrimination among stimuli with common elements involves the ignoring of these common elements. Failure to sample selectively in a discrimination task having common elements leads to errors. Therefore, one expectation from the hypothesis is an interaction between learning ability and matching behavior—to common test cues inferior learners should be more likely to match later (2:1:4:1) frequencies and superior learners earlier frequencies (1:1:1:1). This is discussed more fully in LaBerge and Smith (1957) and Binder and Feldman (1960).

Since the data indicate that if selective sampling occurs it does so by the third block, the data from Groups II and III were combined, as were the data from Groups IV and V. For the inferior learners of Groups II and III the summed observed frequencies are 53 and 18, which is very close to the 3:1 ratio. The comparable frequencies for the superior learners are 40 and 23. This differs significantly from both equal frequency and 3:1 responding ($\chi^2 = 4.58$ and 4.45, respectively, $p \leqslant .05$). However, 40 high and 23 low does not differ significantly from the expectation based on the assumption of overall matching ($\chi^2 = 2.03$).

In the case of Groups IV and V the difference between inferior and superior learners in high to low ratios, 44:24 (1.84:1) and 75:45 (1.67:1), respectively, is too slight to make any further comparisons necessary. And, in fact, in all four of the relevant comparisons, the overall expectation is closer to observation than either 1:1:1:1 or 2:1:4:1.

DISCUSSION

The learning curves (Fig. 2) indicate that there are no systematic differences among groups in the final blocks of trials. The maximum difference at the block designated 14 is only .05 errors per attempted anticipation and all groups seem to have achieved relatively stable response levels. The starting point for Group I of .44 is not surprising since all plotting started with the second actual block for each group and the first block for Group I consisted of eight trials.

The overall picture presented in Table 2 seems to point to the tendency toward using the frequency information over the entire learning process. Considering that a minimum of 10 blocks intervenes between the last equally frequent block and the test, the closeness of the fit between "Overall matching" and observed results is very impressive. Whether this indicates an acute sensitivity of *S*s to relatively fine differences in the relative frequencies of objects in their environments as combined with long storage capacities or merely the lingering impact of initial exposure is not clear.

The findings of Binder and Feldman (1960) and LaBerge and Smith (1957) of interaction between learning ability and matching tendencies in response to a set of background or common cues received only partial support in the present experiment. The differential frequencies for inferior and superior learners observed in Group II are very similar to those found by Binder and Feldman for a group shifted from 1:1:1:1 to 2:1:4:1 after three blocks, as are the observed frequencies for the inferior learners of Group III. The 2:1 (24:12) observed ratio for the superior learners of Group III produced the significant deviation of the total for superior learners from the result previously found.

Binder and Feldman found that superior learners shown C′ and C″ (Fig. 1) came close to matching the differential rate (after three blocks of 1:1:1:1 followed by 2:1:4:1 to criterion—this came to an average of 8.24 differential blocks). They argued that since C′ and C″ contained discriminative cues, *S*s did not ignore them after the shift in frequencies. While the observed frequencies for the C′ and C″ *S*s of the present experiment are not far from those of Binder and Feldman (their high-low ratio was 33:14), there would seem to be grounds for questioning the reasonableness of expecting 2:1:4:1 matching to C′ and C″. The cues of C′ and C″ are more discriminative than the cues of C, but clearly they are insufficient for perfect acquisition. Before an *S* can respond without error he must learn that such combinations of cues (patterns) as triangle and line in bottom right corner are associated with a given name. Perfect learners would then be expected to use these patterns for response determination since any of the cues in C or C′ or C″ would involve a probability less than one of a perfect response. Consequently, at some point during learning, the components of C′ and C″ will be ignored and attention paid to the patterns; this may occur before the experimentally controlled shift to differential frequencies and thus produce a tendency toward 1:1:1:1 matching to C′ and C″.

The data of Table 2 do provide support in that in each of the four groups shown blocks 1:1:1:1 (but not in the control group) the high to low ratio for inferior learners exceeds that for superior learners.

But, one might argue, since they are superior learners perhaps these results mean only that they have better memories and so the 1:1:1:1 blocks have more of a lingering effect. If this were indeed the case, however, one would expect the ratio of high to low frequencies to be distinctly less after four blocks than after two blocks of 1:1:1:1. Clearly no such trend for superior learners (or for inferior learners) is evident in Table 2.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

BINDER, A., & FELDMAN, S. E. The effects of experimentally controlled experience upon recognition responses. *Psychol. Monogr.*, 1960, **74**(9, Whole No. 496).

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

LABERGE, D. L., & SMITH, A. Selective sampling in discrimination learning. *J. exp. Psychol.*, 1957, **54**, 423–430.

(Received December 21, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 560–567

NOVELTY, COMPLEXITY, INCONGRUITY, EXTRINSIC MOTIVATION, AND THE GSR[1]

D. E. BERLYNE, MARGARET A. CRAW, P. H. SALAPATEK,
AND JUDITH L. LEWIS

University of Toronto

Skin resistance was recorded from human Ss, while each item of a sequence of visual patterns received 3 successive 3-sec. exposures at 12-sec. intervals. The patterns comprised "less irregular" and "more irregular" items, representing a number of "complexity" and "incongruity" variables. Extrinsically-motivated Ss, i.e., those who were told to attend carefully because they would later undergo a recognition test, produced more frequent GSRs than Ss without such extrinsic motivation. There was a decline in GSR frequency after the 1st exposure of a pattern but a revival when the next pattern appeared. There was also a long-term decline over patterns. Incongruous pictures evoked GSRs of greater mean amplitude than nonincongruous pictures. Some support was obtained for the conclusion that "more irregular" patterns are more likely to evoke GSRs than "less irregular" patterns.

The experiments to be reported are part of a research program on relations between the orientation reaction and collative properties of stimulus patterns, i.e., properties that are customarily covered by such words as "novelty," "surprisingness," "complexity," and "incongruity" (Berlyne, 1960, 1963b). These have been shown by a number of techniques to have a potent influence over exploratory behavior. They appear also to have other important motivational effects, e.g., in fear, humor, and aesthetics. There are reasons for believing that conflict underlies them, and they are, further, closely related to the concepts handled by information theory.

The orientation reaction (see Berlyne, 1960, 1963b) is a complex of psychophysiological changes that have been shown to accompany at least some forms of exploratory behavior. They are indicative of a rise in arousal or drive. The present experiments concentrated on the GSR as one of the most sensitive components of the orientation reaction, but it must be remembered that skin resistance is, at best, an indirect and imperfect measure of arousal.

The stimulus material used in the present experiments (see Berlyne & Lawrence, in press) is exhibited in Fig. 1. It comprises eight categories, each containing two pairs of visual patterns. Each pair consists of a "less irregular (LI)" pattern (displayed on the left) and a "more irregular (MI)" pattern (displayed on the right), but the variable that distinguishes MI from LI patterns differs from category to category, as the captions denote. These variables are, nevertheless, alike in that (*a*) they are all, with the possible exception of Variable E, variables to which everyday language might well refer as differences in "complexity," (*b*) they appear to have similar effects on exploratory behavior (Berlyne, 1957, 1958, 1960, 1963a; Berlyne & Lawrence, in press; Berlyne & Lewis, in press), (*c*) the MI patterns, according to the theory offered by Berlyne (1960), embody more "arousal potential" than the LI

---

[1] This investigation was supported by Research Grant M-6324 from the National Institute of Mental Health, United States Public Health Service.

patterns, and (*d*) the MI patterns, given certain assumptions, may be regarded as having a higher information content (Attneave, 1954; Garner, 1962). For these reasons, these variables will be discussed collectively as well as singly.

A further opportunity to test the

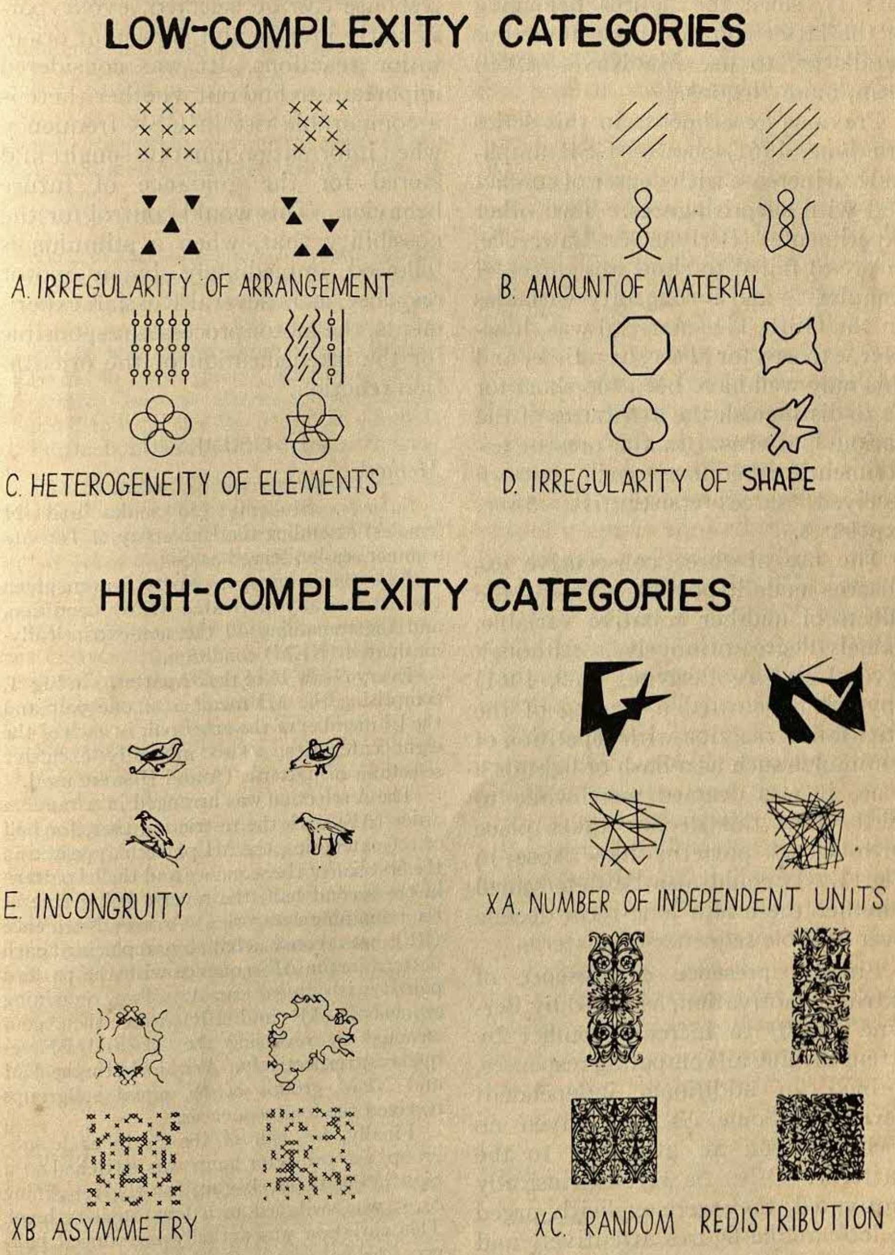

LOW-COMPLEXITY CATEGORIES
A. IRREGULARITY OF ARRANGEMENT
B. AMOUNT OF MATERIAL
C. HETEROGENEITY OF ELEMENTS
D. IRREGULARITY OF SHAPE
HIGH-COMPLEXITY CATEGORIES
E. INCONGRUITY
XA. NUMBER OF INDEPENDENT UNITS
XB ASYMMETRY
XC. RANDOM REDISTRIBUTION

FIG. 1. The stimulus material.

effects of a "complexity" variable is offered by a comparison between Categories A–D and Categories E–XC (headed "low-complexity categories" and "high-complexity categories" in Fig. 1), since the figures belonging to the latter contain more numerous details or, to use MacKay's (1950) term, more "logons."

Previous experiments in this series (Berlyne, 1961) showed GSR amplitude to increase with degree of conflict and with surprisingness. Two other experiments (Berlyne & Lawrence, in press) failed to show any effect of complexity and incongruity variables on the GSR. The material was, however, exposed for .2 sec. at a time, and this may well have been too short for *S*s to distinguish the attributes of the various patterns. In the present experiments, therefore, each pattern received three consecutive 3-sec. exposures.

The use of three consecutive exposures made it possible to assay the effects of another collative variable, namely degree of novelty. Although several *E*s (see Berlyne, 1960, 1961) have demonstrated a waning of the orientation reaction with repetition of a stimulus such as a flash of light or a tone, it was deemed worthwhile to verify that this decline takes place with visual patterns like those in Fig. 1. It could also be determined whether there is a long-term decline over a whole sequence of patterns.

Finally, presence or absence of extrinsic motivation, assumed by Berlyne (1960) to increase conflict by strengthening all competing responses, formed an additional independent variable. Some *S*s were given no special reason for attending to the patterns. Other *S*s were extrinsically motivated: they were explicitly urged to look at the figures attentively and given to understand that they would later undergo a recognition test. Russian *E*s (Sokolov, 1958) have regularly reported that "signal stimuli," i.e., stimuli whose properties must be discriminated so that an overt response can be selected, evoke particularly intense and persistent orientation reactions. It was considered important to find out whether there is a comparable rise in GSR frequency when information must be sought and stored for the guidance of future behavior. This would control for the possibility that, when a stimulus is followed immediately by a motor response, as in several Russian experiments, the motor process is responsible for the intensification of the orientation reaction.

EXPERIMENT I

Method

Subjects.—Students (53 males and 27 females) attending the University of Toronto summer session served as *S*s.

Experimental design.—Forty *S*s were given the extrinsically-motivated (EM) condition, and the remaining 40 the non-extrinsically-motivated (NEM) condition.

Every *S* saw 16 of the 32 patterns in Fig. 1, comprising the MI member of one pair and the LI member of the other pair in each of the eight categories. Two mutually exclusive selections of patterns (A and B) were used.

The A selection was arranged in a random order (AF) with the restriction that, for half of the categories, the MI pattern appeared in the first half of the sequence and the LI pattern in the second half, the reverse being true of the remaining categories. A second sequence (BF) was constructed by replacing each pattern in the AF sequence with the pattern paired with it in Fig. 1. Two remaining sequences (AR and BR) were then constructed by reversing the AF and BF sequences, respectively. Within both the EM and NEM groups of *S*s, equal subgroups received the four sequences.

Finally, for half of the *S*s in each subgroup, the projector lamp was switched to a high intensity level, and, for the remaining *S*s, it was switched to a low intensity level. This variation was actually introduced halfway through the procedure as an economy measure, in order to conserve bulbs.

There was thus a 2 × 4 × 2 design, with five *S*s in each of the eight cells.

Apparatus.—Skin resistance was measured by means of a Grass Model 5 polygraph, placed in an adjoining room, with Fels mercury-coated zinc electrodes attached to both palms. An automatic projector exposed patterns on a screen for 3 sec. at a time, with an interval of 12 sec. between exposures.

Procedure.—The electrodes were placed on *S*'s palms and he was made to wait for 20–30 min., so that hydration could be completed. He was then led into the experimental room and seated in an easy chair in front of the screen.

The *S*s of the EM group were told:

> Some pictures are going to appear on the screen while certain physiological measures are taken. The first five pictures will appear once each, but after that, each picture will appear three times. You are asked to pay special attention to the pictures, because afterwards you will be asked to recognize the ones that you have seen among others that are similar to them.

The *S*s in the NEM group were given the identical instructions except that, in place of the last sentence, they were told: "You will not be asked any questions about the pictures at any time, but please keep your eyes fixed on the screen."

The light was then turned out, and *E* left the experimental room. Five practice patterns were shown once each. They resembled the LI patterns of Categories A–E, and their function was to decrease variability by absorbing the marked decline in GSR amplitude that invariably occurs during the first few exposures.

Immediately after the 5 practice patterns, the 16 experimental patterns appeared three times each. After the last exposure of the fourth experimental pattern, and again after the last exposure of the tenth, *E* went back into the experimental room and turned on the light. This was partly to change the slide tray on the projector and partly to rouse *S* and stave off habituation. If GSRs had ceased to appear by this point, as was not infrequently the case, they generally revived after the interruptions.

Results

The mean number of GSRs was treated as the principal dependent variable. A GSR was counted if (*a*) a downward deflection in the resistance tracing appeared after the onset of an exposure and no later than 4 sec. after its termination, and (*b*) a downward deflection was followed within 10 sec. by a trough and subsequent return upwards.

Intensity of illumination.—No significant differences distinguished *S*s that had the projector lamp at higher and lower intensities.

Extrinsic motivation.—Disregarding the five preliminary practice patterns, each *S* had 48 opportunities to show a GSR. The mean number of GSRs was significantly higher in the EM group than in the NEM group (see Table 1), $t\ (78) = 3.42, p < .001$.

Novelty.—As shown in Table 1, the mean number of GSRs declined from the first to the second presentation of a pattern, with no marked difference between the second and third presentations. Friedman ranking analyses of variance showed the means for the three presentations to differ significantly. For the EM group, $\chi_r^2\ (2) = 16.66,\ p < .001$. For the NEM group, $\chi_r^2\ (2) = 7.36,\ p < .05$. For both groups together, $\chi_r^2\ (2) = 19.95,\ p < .001$.

This result is, however, compatible with either a steady decline in the number of GSRs from the beginning to the end of the experimental session or a "saw-tooth" pattern, such that the probability of a GSR decreased

TABLE 1

TEMPORAL-ORDER EFFECT BY PRESENTATIONS: EXP. I

Group	Presentation			
	I	II	III	Total
EM	9.8	7.5	8.5	25.8
NEM	6.4	5.3	4.7	16.4
Average	8.1	6.4	6.6	21.1

Note.—Mean no. GSRs.

with successive presentations of one pattern but revived at the first presentation of a new pattern. In order to decide between these possibilities, it is necessary to compare the mean number of GSRs for the third presentation of every pattern that was followed by another pattern and for the first presentation of every pattern that was preceded by another pattern. For the EM group, the mean number of GSRs rose from 7.9 for the last presentation of a preceding pattern to 9.0 for the first presentation of a following pattern, t (39) = 2.73, $p < .01$. For the NEM group, the corresponding means were 4.6 and 5.7, t (39) = 1.94, $.05 < p < .10$. For both groups together, the means were 6.3 and 7.4, t (79) = 3.65, $p < .001$.

Effects of temporal order on the mean number of GSRs over all three presentations of a pattern were then examined. The design of the experiments ensured, of course, that temporal order was orthogonal to all the other comparisons to be made. The mean numbers of GSRs (over three

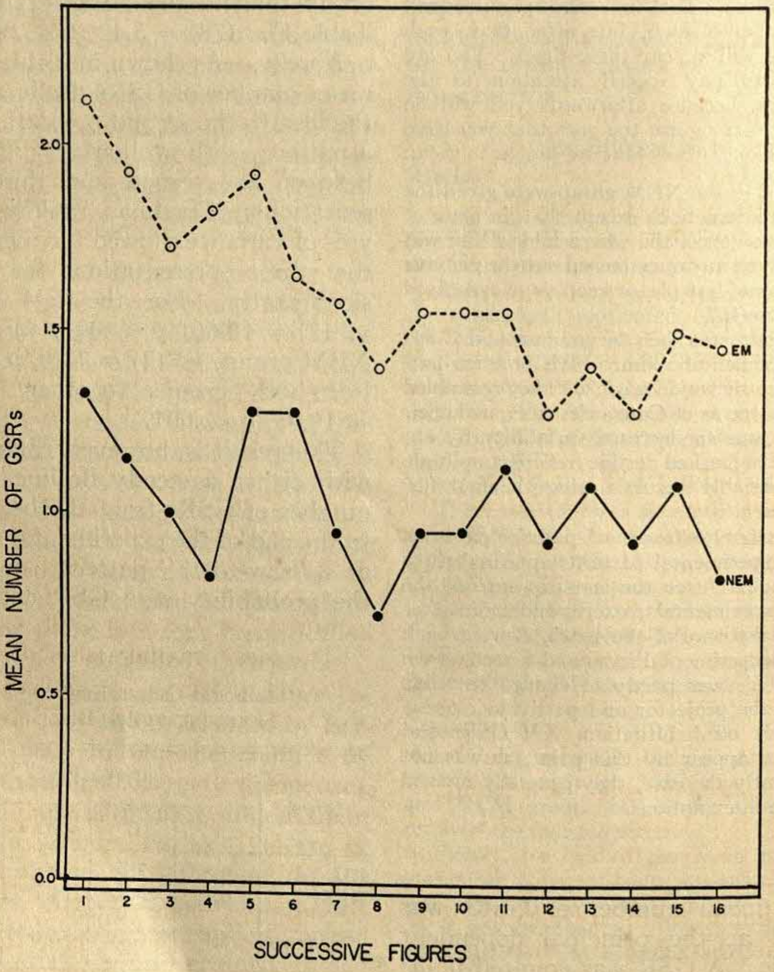


FIG. 2. Temporal-order effects by patterns: Exp. I.

TABLE 2

EFFECTS OF COMPLEXITY AND INCONGRUITY VARIABLES: EXP. I

Categories	NEM Group			EM Group		
	LI Patterns	MI Patterns	All Patterns	LI Patterns	MI Patterns	All Patterns
A	0.9	1.0		1.6	1.7	
B	1.1	1.0		1.5	1.6	
C	1.1	1.1		1.5	1.6	
D	1.0	1.0		1.4	1.6	
E	1.0	1.0		1.6	1.9	
XA	1.0	1.1		1.4	1.5	
XB	1.0	0.9		1.8	1.8	
XC	1.2	1.0		1.5	1.7	
A–D	4.1	4.2	8.3	6.0	6.5	12.5
E–XC	4.1	4.0	8.1	6.4	6.8	13.2
All categories	8.2	8.2	16.4	12.4	13.3	25.7

Note.—Mean no. GSRs rounded off to one decimal place.

presentations) for the patterns that appeared first, second, third, etc. are displayed in Fig. 2. For both EM and NEM groups, there is a visible decline, which is most marked at the beginning of the sequence, even though five practice patterns had already appeared. To test the significance of this decline, the total numbers of GSRs per pattern were transformed into ranks, and the ranking correlation coefficient (Kendall's τ) between the rankings obtained in this way and the temporal order was calculated. The value of τ was $+.72$ ($p < .001$) for the EM group, $+.33$ ($p < .05$) for the NEM group, and $+.55$ ($p < .01$) for both groups together.

Complexity variables.—The means that are necessary for assessing the effects of complexity variables are shown in Table 2. In the NEM group, there is no sign of a significant difference between MI and LI patterns or between high-complexity and low-complexity categories. On the other hand, there are sizable differences in the predicted directions for the EM group. Analysis of variance of the data for the EM group reveals that, for the MI/LI comparison, $F\ (1, 597) = 3.83$, $p < .026$. As for the comprison between high-complexity and low-complexity categories, $F\ (1, 597) = 1.62$, which is not significant.

The significance of the interaction between the EM/NEM and MI/LI variables was tested as follows. The number of GSRs to LI patterns was subtracted from the number of GSRs to MI patterns for each S, and the Mann-Whitney test was used to compare these difference scores for the EM and NEM groups. There was evidence of a significant interaction: $U = 212.5$, $p < .05$.

The mean numbers of GSRs for LI and MI patterns by categories are displayed in Table 2. It will be noted that no important differences are in evidence in the NEM group. But in the EM group, the MI mean is higher than the LI mean for seven categories out of eight with one tie; $p < .01$ according to the sign test.

Incongruity.—The data are not numerous and sensitive enough to permit a meaningful comparison of mean numbers of GSRs for MI and LI patterns within categories. However, an amplitude score was worked out for every GSR, using the log-conductance-change measure recommended by Haggard (1949), i.e.,

$$\log_{10}\left[10^9\left(\frac{1}{R_m}-\frac{1}{R_i}\right)+1\right],$$

where R_i is the resistance at the initial deflection point and R_m the resistance at the trough, both measured in ohms. The mean amplitude scores for the first presentation of the MI and LI figures of each category were taken, omitting those *S*s in whom neither figure in that category produced a GSR.

None of these within-category differences was significant with the exception of those in Category E (incongruity). In this category, the mean magnitude scores in the EM group were 1.86 for the MI pattern and 1.23 for the LI pattern, counting 34 *S*s, $t\,(33) = 2.57$, $p < .01$. The means for the NEM group, counting 27 *S*s, were 1.48 and 0.98, $t\,(26) = 1.70$, $p < .05$, (one-tailed). For both groups together, the means were 1.69 and 1.12, $t = 2.93$, $p < .01$.

EXPERIMENT II

Experiment II was undertaken with a view to obtaining more data on the effects of the complexity variables.

Method

The design and procedure were identical with those used for the EM group of Exp. I.

Forty undergraduates (20 males and 20 females) from a first-year introductory psychology class served as *S*s.

Results

The only data analyzed were the mean numbers of GSRs relevant to the effects of complexity variables. They are shown in Table 3.

TABLE 3

EFFECTS OF COMPLEXITY AND INCONGRUITY VARIABLES: EXP. II

Categories	Patterns		
	LI	MI	All
A	1.3	1.2	
B	1.4	1.2	
C	1.1	1.2	
D	1.3	1.2	
E	1.2	1.4	
XA	1.2	1.3	
XB	1.5	1.2	
XC	1.3	1.3	
A–D	5.2	4.8	10.0
E–XC	5.2	5.3	10.5
All categories	10.4	10.1	20.5

Note.—Mean no. GSRs rounded off to one decimal place; extrinsically-motivated condition—40 *S*s.

The overall mean number of GSRs for this experiment fell about midway between the means for the EM and NEM groups of Exp. I. There is, therefore, reason to suspect that, for some reason, these *S*s had less overall attentiveness than the EM group of Exp. I.

In Exp. II, the excess of GSRs to MI patterns over GSRs to LI patterns did not appear. There was actually a slight difference in the opposite direction. There was, however, a difference in the predicted direction between high-complexity and low-complexity categories.

Analysis of variance showed neither of these differences to be significant in Exp. II alone. For the MI/LI comparison, $F\,(1, 597) = 0.45$, and for the high-complexity-category/low-complexity-category comparison, $F\,(1, 597) = 1.34$. When, however, the significance levels for the high-complexity-category/low-complexity-category comparison obtained in Exp. II and in the EM group of Exp. I are combined, using the Stouffer technique (Mosteller & Bush, 1954), $z = 1.72, p < .05$. If we examine all four F tests relevant to complexity variables in the two groups, $z = 1.86$, and $p \sim .03$.

DISCUSSION

There is thus some indication of a greater incidence of GSRs with more complex or incongruous visual patterns, but this effect is only in evidence when *S*s are highly attentive. And the effects of complexity variables on the GSR are certainly not comparable with the pronounced effects that these same variables have on exploratory behavior.

It may, however, be concluded with reasonable confidence that GSR incidence increases with novelty (as shown by the decline over successive patterns and the decline over successive presentations of one pattern), that GSR incidence increases with extrinsically motivating instructions, and that GSR amplitude increases with incongruity.

REFERENCES

ATTNEAVE, F. D. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61,** 183–193.

BERLYNE, D. E. Conflict and information-theory variables as determinants of human perceptual curiosity. *J. exp. Psychol.*, 1957, **53,** 399–404.

BERLYNE, D. E. The influence of complexity and novelty in visual figures on orienting responses. *J. exp. Psychol.*, 1958, **55,** 289–296.

BERLYNE, D. E. *Conflict, arousal and curiosity.* New York: McGraw-Hill, 1960.

BERLYNE, D. E. Conflict and the orientation reaction. *J. exp. Psychol.*, 1961, **62,** 476–483.

BERLYNE, D. E. Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. *Canad. J. Psychol.*, 1963, **17,** 274–290. (a)

BERLYNE, D. E. Motivational problems raised by exploratory and epistemic behavior. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 5. New York: McGraw-Hill, 1963. (b)

BERLYNE, D. E., & LAWRENCE, G. H. Effects of complexity and incongruity variables on GSR, investigatory behavior and verbally expressed preference. *J. gen. Psychol.*, in press.

BERLYNE, D. E., & LEWIS, J. L. Effects of heightened arousal on exploratory behavior. *Canad. J. Psychol.*, in press.

GARNER, W. R. *Uncertainty and conflict as psychological concepts.* New York: Wiley, 1962.

HAGGARD, E. A. On the application of analysis of variance to GSR data: I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, **39,** 378–392.

MCKAY, D. M. Quantal aspects of scientific information. *Phil. Mag.*, 1950, **41,** 289–311.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954.

SOKOLOV, E. N. *Vospriiate i uslovny refleks.* [Perception and the conditioned reflex.] Moscow: Moscow Univer. Press, 1958.

(Received December 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 568–571

LEFT-RIGHT DIFFERENCES IN TACHISTOSCOPIC RECOGNITION[1]

M. P. BRYDEN[2] AND CHRISTOPHER A. RAINEY[3]

McGill University

3 types of material, letters, geometric forms, and outline drawings of familiar objects, were presented tachistoscopically to 32 Ss. In 1 series, the figures appeared successively in either the left or the right visual field; in a 2nd series, 2 figures appeared simultaneously, 1 in each field. In the simultaneous presentation condition, all types of material were more readily recognized in the left visual field, a phenomenon which appears related to order of report. With successive presentation, a right-field superiority, greatest for letters and least for geometric forms, was observed. It is suggested that the left-right differences in successive presentation increase with increasing stimulus familiarity. Emphasizing fixation instructions decreased overall accuracy and accentuated left-right differences in both presentation conditions.

A number of studies have demonstrated differences in the ease of identifying material exposed tachistoscopically in the left visual field and that exposed in the right visual field. Both the method of presenting the stimuli and the type of material used have been found to influence left-right differences.

When stimuli are presented in both visual fields simultaneously accuracy is generally superior in the left visual field. Heron (1957) has observed this with letters, and Bryden (1960) with both letters and geometric forms. In contrast to these reports, however, Wyke and Ettlinger (1961) have found a right-field superiority for outline drawings of familiar objects exposed in both visual fields at the same time.

Quite different results have been found when stimuli are presented successively in one visual field or the other. Superior recognition of material exposed in the right visual field has been observed with English words (Mishkin & Forgays, 1952; Terrace, 1959), letters (Heron, 1957), and outline drawings of familiar objects (Wyke & Ettlinger, 1961). On the other hand, both geometric forms (Bryden, 1960; Heron, 1957) and nonsense forms (Heron, 1957; Terrace, 1959) are recognized with equal facility in both visual fields.

It is not entirely clear from the above studies when left-right differences in accuracy should be expected, nor is it clear what factors produce these differences. The present study was designed to investigate more fully the conditions in which left-right differences in tachistoscopic recognition occur.

METHOD

Three different types of material, capital letters, simple geometric forms, and drawings of familiar objects, were used in this experiment. Each of these types of material was presented in two ways: successively in either the left or the right visual field, and simultaneously in both visual fields. Finally, since there was some question as to whether Ss were actually looking at the pre-exposure fixation point, two different types of fixation instruction were used, one of which emphasized the importance of fixation while the other did not.

---

[1] This research was supported by Grant No. 9401-11 from the Defence Research Board of Canada to D. O. Hebb.

[2] Now at the University of Waterloo, Ontario.

[3] Now at Queens University of Belfast, Northern Ireland.

Apparatus and materials.—The stimulus elements used were eight capital letters, eight outline geometric forms (square, triangle, star, etc.), and eight outline drawings of familiar objects such as a fish, a lamp, a face, and a TV set. These figures were drawn in India ink on 3 × 5 in. white cards, and filled an area 1 in. square. The figures were drawn by tracing from a standard, with lines about $\frac{1}{32}$ in. wide. The cards for successive presentation each contained a single figure, either to the right or to the left of the center; while those for simultaneous presentation contained two different figures of the same type, one on either side. Every figure appeared equally often on the left and on the right in each condition. A total of 48 different cards was prepared for each presentation condition.

The cards were presented with a Gerbrands mirror tachistoscope, set for an exposure time of 10 msec. When exposed, each figure was centered 2°52′ to one side of a central fixation point, and subtended a visual angle of 4°17′. The central fixation point was visible in the pre-exposure field.

Subjects.—The *S*s were 32 college undergraduates; none had any serious uncorrected visual defects.

Procedure.—Half the *S*s were shown the simultaneous presentation material first, and half saw the successive material first. Within a particular presentation condition, the cards appeared in a completely random order, so that *S* did not know which type of material would appear, nor, in the successive condition, where the figure would appear.

All *S*s were instructed at the beginning of the experiment to fixate the central dot prior to every exposure. For 16 of the *S*s, this point was not emphasized again (weak instructions). The remaining 16 *S*s were reminded to fixate the central dot prior to each exposure (strong instructions). After each presentation, *S* reported what he saw orally to *E*.

Scoring.—A correct response was scored when *S* identified the material properly. Some of the familiar objects were frequently assigned a plausible, but erroneous, label; these were also scored as correct. While *E* was aware of the conditions under which the erroneous form responses were obtained, only the responses used several times for a single item were scored as correct.

RESULTS

In general, a left-field superiority was observed with simultaneous presentation, and a right-field superiority was found with the successive presentation procedure. The degree of left- or right-field superiority in the various conditions is shown in Fig. 1. In this figure, a positive score indicates a right-field superiority.

Mean recognition scores for each condition are shown in Table 1. Preliminary analyses of variance indicated that the order in which the two presentation conditions were administered had no significant effect; this factor is therefore not included in the final analyses.

An analysis of variance for the simultaneous presentation condition is summarized in Table 2. In this condition, all types of material were more readily identified in the left visual field. Since the Location

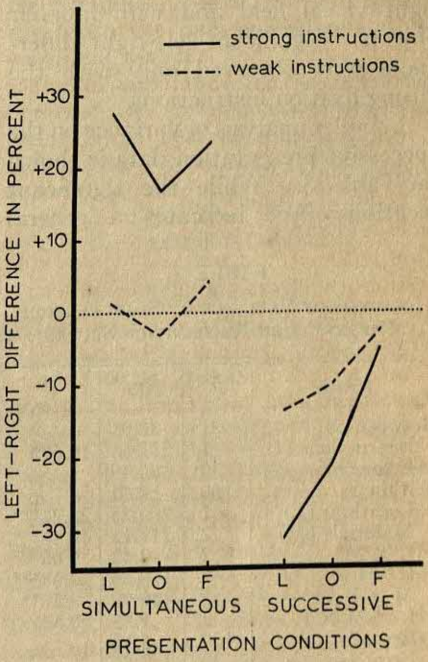


FIG. 1. Visual field differences as a function of method of presentation, type of material, and instructions. (Scores represent differences between the two fields in accuracy; a positive score indicates left-field superiority, a negative score indicates right-field superiority.)

TABLE 1

Percent Recognition Scores for Left and Right Visual Fields

Presentation Cond.	Instructions	Type of Material							
		Letters		Objects		Forms		Overall	
		L	R	L	R	L	R	L	R
Simultaneous	Weak	71.9	70.6	62.5	65.6	80.6	76.3	71.7	70.8
	Strong	71.9	44.4	50.6	33.8	82.5	58.9	68.3	45.7
	Average	71.9	57.5	56.5	49.7	81.6	67.5	70.0	58.2
Successive	Weak	53.8	67.5	53.8	63.8	65.0	67.5	57.5	66.3
	Strong	32.5	63.8	32.5	53.8	70.0	75.0	45.0	64.2
	Average	43.1	65.6	43.1	58.8	67.5	71.3	51.2	65.2

× Material interaction is not significant, the degree of left-field superiority is similar for all types of material. Accuracy was much lower with the strong instructions than with the weak instructions, especially in the right visual field and with familiar objects. Significant left-right differences were observed only with the strong fixation instructions.

A similar analysis of variance on the successive presentation data is shown in Table 3. While the significant location effect indicates a general right-field superiority, in this case the Location × Material interaction is significant. An examination of Fig. 1 shows that the right-field superiority is greatest for letters and least for geometric forms; *t* tests indicate that accuracy is significantly better in the right visual field only for letters and familiar objects. The degree of right-field superiority is significantly greater for letters than for geometric forms ($p < .01$), greater for letters than for familiar objects ($p < .10$),

TABLE 2

Analysis of Variance on Simultaneous Presentation Recognition Scores

Source	*df*	*MS*	*F*
Between *S*s	31	38.94	
Instructions (I)	1	292.55	19.59*
Error (b)	30	30.49	
Within *S*s	160	10.41	
Locations (L)	1	206.25	26.24***
Materials (M)	2	217.14	40.44***
L × M	2	2.42	0.66
L × I	1	178.26	22.68***
M × I	2	30.99	5.77**
L × M × I	2	1.88	0.52
Error (w)	150	5.18	
Error (L)	30	7.86	
Error (M)	60	5.37	
Error (L × M)	60	3.65	
Total	191	15.04	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 3

Analysis of Variance on Successive Presentation Recognition Scores

Source	*df*	*MS*	*F*
Between *S*s	31	6.41	
Instructions (I)	1	15.19	2.48
Error (b)	30	6.12	
Within *S*s	160	2.63	
Locations (L)	1	60.75	39.99***
Materials (M)	2	39.20	20.99***
L × M	2	8.73	7.59**
L × I	1	8.33	5.49*
M × I	2	13.08	7.00**
L × M × I	2	1.44	1.25
Error (w)	150	1.51	
Error (L)	30	1.52	
Error (M)	60	1.87	
Error (L × M)	60	1.15	
Total	191	3.24	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

and greater for objects than for forms ($p < .05$).

Fixation instructions also had a considerable effect on successive recognition scores. The left-right differences were exaggerated by the strong instructions, primarily due to a drop in left-field accuracy. Once again, recognition of the familiar objects was most influenced by the instructions.

DISCUSSION

When material is exposed simultaneously in both visual fields, recognition is superior in the left visual field. While similar results have been reported previously for both letters and geometric forms, the present findings are in marked contrast to those of Wyke and Ettlinger (1961) with familiar objects. The failure to confirm Wyke and Ettlinger's findings is most likely due to procedural differences. They used a rather novel limits procedure for assessing accuracy, and presented a series consisting solely of outline drawings of familiar objects, rather than employing several different types of material.

Observing that most *S*s report simultaneously-presented material in a left-to-right order, Bryden (1960) suggested a fading-trace theory to account for the left-field superiority. There is more time between presentation and report for the trace of the figure on the right to fade below response threshold, and thus it would be identified correctly less often. This notion would seem to handle the present data quite adequately.

With successive presentation, the degree of right-field superiority varies with the type of material. The greatest left-right difference is shown with letters, while there is no significant right-field superiority with geometric forms. It is plausible that left-right differences are related in some way to familiarity with the material. Without question, the letter material is the most familiar. The outline drawings of familiar objects are not themselves familiar, but the objects they represent are very common, while the geometric forms are relatively rare. There thus seems to be some logical support for the idea that right-superiority covaries with stimulus familiarity. If this notion is tenable, the degree of right-field superiority found with nonsense forms should be less than that for geometric forms. In fact, Heron (1957) has found a smaller right-superiority with nonsense forms than with geometric forms, and Terrace (1959) reported an almost significant left-superiority with nonsense forms.

Kimura (1961) found a relation between right-superiority in certain auditory tasks and speech localization, and suggested that a similar relation might be found with tachistoscopic tasks. She argued that the dominant hemisphere is involved in the identification of verbally-related material. Thus verbal materials should be better identified on the right and nonverbal materials on the left. While the present experiment provides no concrete evidence relevant to this suggestion the data would seem to be consistent with it.

In conclusion, then, the left-superiority observed with simultaneous presentation seems to be best accounted for by a fading-trace hypothesis. The right-superiority found with successive presentation, on the other hand, appears to be related to stimulus familiarity.

REFERENCES

BRYDEN, M. P. Tachistoscopic recognition of non-alphabetical material. *Canad. J. Psychol.*, 1960, **14**, 78–86.

HERON, W. Perception as a function of retinal locus and attention. *Amer. J. Psychol.*, 1957, **70**, 38–48.

KIMURA, D. Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychol.*, 1961, **15**, 166–171.

MISHKIN, M., & FORGAYS, D. G. Word recognition as a function of retinal locus. *J. exp. Psychol.*, 1952, **43**, 43–48.

TERRACE, H. S. The effects of retinal locus and attention on the perception of words. *J. exp. Psychol.*, 1959, **58**, 382–385.

WYKE, M., & ETTLINGER, G. Efficiency of recognition in left and right visual fields: Its relation to the phenomenon of visual extinction. *Arch. Neurol., Chicago*, 1961, **5**, 659–665.

(Received December 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 572–579

RELATION BETWEEN PHYSIOLOGICAL RESPONSES TO ENVIRONMENTAL HEAT AND TIME JUDGMENTS

C. R. BELL AND K. A. PROVINS[1]

Medical Research Council Unit, London School of Hygiene and Tropical Medicine

In 3 experiments changes in the magnitude of estimations of standard intervals of 45 sec., 100 sec., 4 min., 9 min., and 13 min. were examined when body temperature and pulse rates had been increased by exposure to high temperature environmental conditions. 3 methods of estimation were used: counting at an estimated rate of 1 digit per second, verbal estimation of time passed in a given activity, and the productions of an interval named, but not demonstrated by E. No consistent relationship was found between changes in estimate size and changes in either body temperature or pulse rate. Wide inter-S variation was found in both direction and extent of change in estimate size with significant increases in body temperature and pulse rate.

Francois (1927) has reported a study in which three *S*s were required to tap a Morse key at an estimated rate of three taps per second for four periods each of 10 sec. When axilla temperature was raised by diathermy (a high frequency alternating current passed through the body) the speed of tapping increased. Hoagland (1933) similarly found that subjectively counting 60 sec. took less time when influenzal fever (two *S*s) or diathermy (one *S*) had raised body temperature.

Hoagland (1935) had found in a series of experiments that for poikilothermic organisms (i.e., those not possessing temperature controlling mechanisms) there was a systematic relationship between the log frequency of events such as heart rates, respiratory movements, beating of the cilia, sensory reaction times, and cricket chirping, and the temperature-sensitive chemical reactions which underlay them. When data from the time estimation studies of Hoagland (1935) and Francois (1927) which related changes in time estimation to changes in body temperature, were examined in a similar way and found to be related in accordance with the general, event frequency/increase in temperature, equation (the Arrhenius equation), Hoagland (1935) concluded that

> psychological time thus seems to depend directly upon the velocities of certain definite chemical processes. . . . Measurements of the estimations of short durations as a function of temperature indicate the existence of a master chemical reaction [pp. 119–120].

This view has been propagated by reviewers of time estimation studies (vide, Pieron, 1945; Woodrow, 1951). More recently, however, Hoagland (1951), Kastenbaum (1959), and Rashkis and Rashkis (1962) have cautioned against the view that time experiences can be adequately explained on the basis of a unitary chemical, temperature-dependent, pacemaker.

Neither Francois (1927), Hoagland (1933), nor Hoagland and Perkins (1935) found evidence for similar changes in time estimation with increases in pulse rate. Schaefer and Gilliland (1938) raised the pulse rates of four *S*s by the vigorous exercise of running up and down steps and found no concomitant changes in time estimation.

---

[1] The authors wish to express their gratitude to the Royal Naval Personnel Research Committee for facilitating Exp. III of this series.

In view of the limitations of the original studies in terms of the small number of *S*s used, the rather special ways in which body temperature was raised, and the fact that 2 of Hoagland's *S*s were not in normal health, it was felt that confirmatory evidence from a larger number of *S*s investigated under laboratory conditions was needed. In the present study the hypotheses were tested that: (*a*) an increase in body temperature would be associated with changes in time estimation in predictable directions; and (*b*) increases in pulse rate would not be associated with changes in time estimation. Three experiments have been carried out, in all of which body (oral) temperature and pulse rate were raised by exposure to high temperature environmental conditions. In Exp. I, 10 *S*s counted at an estimated rate of 1 per second to estimate intervals of 45 and 100 sec. In Exp. II (19 *S*s) and Exp. III (8 *S*s), standard intervals of 4, 9, and 13 min. were estimated by the verbal estimation of time passed method and by the production of a named interval method (Clausen, 1950) in the two experiments, respectively. Relatively long standard intervals were incorporated in the present study because it was felt that these would have more relevance to realistic situations in which time judgments have to be made than the standard intervals of 60 sec. or less most frequently found in published reports (vide, Wallace & Rabin, 1960; Woodrow, 1951).

EXPERIMENT I

Method

Procedure.—Ten healthy unacclimatized male volunteer *S*s, whose mean age was 22.4 yr. (range 18–27 yr.), were required to count from 0 to 45 and from 0 to 100 at an estimated rate of one digit per second. During the estimations *S*s were seated at rest after spending 30 min. resting and completing a personality inventory and a further 20 min. performing a paper-and-pencil task. Each *S* came to the laboratory on two occasions 1 wk. apart. On each day an *S* spent 1½ hr. in a control environment of 20/15° C. (dry bulb/wet bulb) and 1½ hr. stripped to the waist in an environment of 47/35° C. (dry bulb/wet bulb) with an air movement of 25 cm/sec. Only one estimation was made in each environment so that the hot vs. cool comparisons of estimates of each of the standard intervals were made between two estimations 1 wk. apart. The order of climatic exposure and the order of estimation of the two standard intervals conformed to a balanced arrangement over the 10 *S*s of the experiment.

Data.—The time taken to count the estimated 45 and 100 sec. was recorded to the nearest second. The *S*s were given no knowledge of results. Oral temperatures were recorded at frequent intervals from an indwelling sublingual thermocouple. In this experiment no pulse counts were made. At the end of each *S*'s second visit to the laboratory he was asked the question: "Do you think that being in heat makes you count faster or slower?"

Results

Taking as an index of performance the signed discrepancy (i.e., + or −) between the actual time spent in counting (the estimate) and the standard interval itself, an analysis of variance was carried out to determine the inter-*S* variation, the variance from the between-climatic conditions source, and the variance from a comparison of estimates of the 45- and 100-sec. standards. In order to examine this last factor, the 100-sec. estimate, standard discrepancies were multiplied by 45/100 to make the two standards proportionately equivalent. It was found that only the between-standards variance attains statistical significance with F (1, 9) = 5.45, $p < .05$. This suggests the advisability of treating the 45-sec. and 100-sec. data separately.

Although no significant contribution to the estimate data variance was made by the between-climatic conditions comparison, the oral temperatures during estimations in the two conditions were highly significantly different. For the 45-sec. estimations, the mean oral temperature of the 10 *S*s rose significantly ($t = 28.37$, $p < .001$) from 36.76° C. in cool conditions to 38.47° C. in hot conditions. A similar change in oral temperature during 100-sec. estimations from 36.72° C. in cool to 38.44° C. in heat was also highly significant ($t = 18.37$, $p < .001$). Following Hoagland (1933), it was expected that raised body temperatures would lead to a *decrease* in the time taken to count the standard number of seconds. However, although the mean estimates of the group fell from 34.6 sec. to 32.2 sec. for the 45-sec. standard and from 86.3 sec. to 73.7 sec. for the 100-sec. standard with the significant increases in body temperature, of the 10 *S*s in the experiment, 4 *S*s showed an *increased* time to count 45 sec. and 2 *S*s an *increased* time to count an estimated 100 sec. at the higher body temperatures.

To examine the relationship between increase in body temperature and decrease in estimate size, product-moment correlation coefficients were calculated for the 45- and 100-sec. data, respectively. The correlations of $r = -0.34$ ($N = 10$) and $r = +0.21$ ($N = 10$) did not differ significantly from zero, nor did paired-comparison t tests between estimates at the lower and higher body temperatures reveal any significant change in estimate size. There is thus no satisfactory evidence to support the hypothesis that increased body (oral) temperature significantly affects the time taken to count an estimated 45 or 100 sec.

In reply to the question of whether *S*s thought that they had counted more quickly or more slowly in heat, one *S* thought that heat made no difference to his speed of counting; four *S*s thought that heat made them count more slowly; and five *S*s thought that they tended to count more quickly in heat. Of the five *S*s who had in fact counted more slowly in heat and had thus taken longer to count the standard number of seconds, three were of the opinion that they had counted more quickly in heat.

EXPERIMENT II

Method

Procedure.—Nineteen healthy unacclimatized male volunteer *S*s, whose mean age was 29.9 yr. (range 19–47 yr.), participated in an experiment which involved performance on a five-choice serial reaction time task (Leonard, 1959) in climatic conditions of 20/15° C. (dry bulb/wet bulb) and 40/35° C. (dry bulb/wet bulb) with an air movement of 25 cm/sec. In cool conditions *S*s wore normal indoor clothing and in the higher temperature conditions they were stripped to the waist. Total exposure time to each of these environments was approximately $2\frac{3}{4}$ hr. containing four work periods on the five-choice serial reaction time task each lasting 20 min. In each work period of each exposure *S*s were asked the question: "How long is it since this work period began?" This question was put to *S* after either 4, 9, or 13 min. The *S*s were told to call out their answer immediately and to make their estimates as precise as possible. The order of climatic exposure and the order of presentation of the three standard intervals were balanced over the entire experimental series.

Data.—Verbal estimates were recorded by *E* in an antechamber as an *S* called out his response. Oral temperatures were recorded from a clinical thermometer before and after each work period within each exposure to the high temperature conditions, and at the beginning and end of the exposures to the cool climatic conditions. The oral temperature associated with each verbal estimation of each of the three standard intervals was taken as the mean oral temperature for the work period in which the estimation was made. A similar procedure was adopted with respect to pulse rates which were taken concurrently with measurements of oral temperature by *E.*

Results

Not every *S* verbally estimated each of the three standard intervals in both climatic conditions. In order, therefore, to take due account of between-*S*s variation, analyses of variance were carried out on the signed discrepancies (i.e., + or −) between the verbal estimates and standards of the 6 *S*s who estimated both the 4- and 9-min. standards, the 7 *S*s who estimated both the 4- and 13-min. standards, and the 10 *S*s who estimated both the 9- and 13-min. standards in both cool and heat conditions. The signed discrepancies were corrected for equivalence before each analysis was carried out by multiplying the 13-min. discrepancies by a factor of 4/13 or 9/13 and the 9-min. discrepancies by a factor of 4/9. These analyses revealed marked between-*S*s variability ($p < .01$) and significant variance from between-standards ($p < .01$) comparisons. Climatic conditions variance made no significant contribution to the variance in the estimate data except in the 9 vs. 13 min. data analysis ($p < .05$). The verbal estimations of each of the three standards were therefore treated separately and the oral temperatures and pulse rates associated with these particular estimations were compared between the two climatic conditions. Paired-comparison *t* tests between the cool and heat conditions revealed highly significant increases in pulse rates ($p < .001$) and in oral temperatures ($p < .001$), with mean values of +35 beats/min and +1.09° C., respectively.

In the cool conditions mean verbal estimates for the three standard intervals were 4.68 min., 7.44 min., and 11.45 min., and in heat these changed to 4.50 min., 8.56 min., and 13.30 min., respectively. Only the change in the verbal estimates of the 13-min. standard interval attained statistical significance ($t = 2.20$, $p < .05$). Product-moment correlations between the extent of increase in verbal estimate size and increases in pulse rate and oral temperature in no case produced coefficients deviating significantly from zero. In fact, for estimates of the 4-, 9-, and 13-min. standards, respectively, only 2 *S*s in 10, 8 *S*s in 12, and 6 *S*s in 13 did increase their estimate when their body temperatures had been increased. It would therefore seem that there is little support for the hypothesis that increases in oral temperature would be significantly associated with increases in the size of verbal estimates of the three standard intervals.

EXPERIMENT III

Method

Procedure.—Eight unacclimatized healthy male volunteer *S*s, whose mean age was 26.1 yr. (range 21–31 yr.), performed a modified version of a 20-dial watchkeeping task (Broadbent, 1954) in five different climatic conditions: viz. 29.5/24.5° C., 43/35° C., 51/37° C., 51/43° C., and 63/47° C. (dry bulb/wet bulb) in a constant air movement of 25 cm/sec. The *S*s wore denim boiler suits throughout the experiment. The experiments were designed to assess tolerance times to these climates and *S*s were removed from these climates as they neared collapse. Mean times of exposure of the eight *S*s decreased with increasing climatic severity from 240 min. to 19 min. (Provins, Hellon, Bell, & Hiorns, 1962). Climates were presented in order of increasing severity throughout 1 wk., and the same routine was repeated with the same *S*s for a second week.

During the exposures, in which *S* generally remained seated before the 20-dial display, production estimates were made of one, two, or all three standard intervals of 4, 9, and 13 min. No standard was estimated more than once in any exposure and the order of presentation of the standards was randomized over the entire experiment. The production method of estimation was conducted as follows: a green light in the center of the 20-dial display indicated to *S* that a time estimation was to be started as soon as possible. The *S* then pressed a button switch on the left of his chair. This illuminated an indicator showing the period of time to be estimated which remained on display throughout the estimation. Pressing the button switch on the left of the chair marked the beginning of the estimation and a button switch on the right of the chair was used by *S* to end the estimation. Several minutes elapsed between the end of one estimation and the start of the next. In the briefer exposures it was often not possible for an *S* to estimate more than a single standard interval. Throughout the experiment *S*s were given no knowledge of their results.

Data.—Production estimates of the three standard intervals were recorded in an anteroom from an electronic counter in tenths of a minute. Frequent readings of oral temperature were taken from an indwelling sublingual thermocouple throughout all exposures. Pulse rates, by palpation, were taken at irregular intervals by an *O* who was always present in the climatic chamber with *S*.

Results

Estimates of all three standard intervals were obtained from all eight *S*s only in the 29.5/24.5° C., 43/35° C., and 51/37° C. (dry bulb/wet bulb) climates. An analysis of variance of the production estimate data, in terms of signed discrepancies (i.e., + or −) between standard intervals and production estimates of them corrected for equivalence, by multiplying the estimate discrepancies for the 9- and 13-min. standards by 4/9 and 4/13, respectively, was carried out therefore over these three climatic conditions. Significant sources of variance in the production estimate data were found in the between-*S*s ($p < .001$), between-weeks ($p < .001$), and between-standards ($p < .001$) comparisons. It was therefore necessary to treat separately the data from each week and the data from each of the three standard intervals. No significant variance derived from the between-climatic conditions comparison.

One difficulty in the examination of the production estimate data by climatic conditions is that the randomized order of presentation of the three standard intervals produced situations in which a given standard may have been estimated under less physiological stress in a severe climate than in a less severe one. This would occur when a standard was estimated shortly after entry into the severe climate but shortly before collapse in a less severe climate. To obviate this difficulty, a mean oral temperature and mean pulse rate were calculated for the duration of each of the estimations, irrespective of the climatic conditions to which *S* was exposed at the time.

When the data were examined for a possible relationship between the sizes of the estimates of each standard and body temperature, within Wk. 1 and Wk. 2 of the experiment, it was found that of the six product-moment correlation coefficients produced (*N* varied from 59 to 68) only one deviated significantly from zero. This was in respect of the estimates of the 13-min. standard which tended to *decrease* with increased body temperature ($r = 0.446$, $p < .001$) in Wk. 1.

It was also possible to categorize the data in three ranges of oral temperature. Production estimates made at oral temperatures of $\leq 36.99°$ C., 37.00–37.49° C., and $\geq 37.50°$ C. were compared by paired-comparisons *t* tests for each of the three standard intervals in each of the 2 wk. of the experiment. Mean production estimates within each of these three temperature ranges are given in Table 1. None of the *t* test comparisons attained a satisfactory level of statistical significance except in respect of the estimation of the 4-min. standard in the two lower ranges of oral temperature when, in Wk. 1, the production estimates at 37.00–37.49° C. were significantly *larger* ($t = 3.13$, $p < .05$) than the production estimates at oral temperatures ≤ 36.99° C. These data therefore offer no support for the hypothesis that higher oral temperatures are associated with smaller production estimates.

A similar procedure to that used for the examination of the production estimate data by oral temperature was adopted to investigate the relationship between changes in pulse rate on exposure to high temperature environmental conditions and changes in production time estimation. Estimates within three ranges of pulse rates ≤ 99 beats/min, 100–119 beats/min, and ≥ 120 beats/min were compared within Wk. 1 and Wk. 2 of the experiment. Paired-comparison *t* values and mean production estimates for the three pulse rate categories are shown in Table 2. There was no evidence to support the assumption that an increase in pulse rate would be associated with a decrease in the size of production estimates.

TABLE 1

MEAN AND *t* VALUES (Paired Comparisons) FOR PRODUCTION ESTIMATES BETWEEN ORAL TEMPERATURE CATEGORIES

	≤36.99 C.	37–37.49 C.	≥37.50 C.	*t*
4 min.				
Wk. 1	4.84	5.68	—	3.13*
	4.84	—	5.38	0.73
	—	4.81	4.60	0.05
Wk. 2	6.15	5.64	—	1.24
	6.15	—	5.61	1.76
	—	5.64	5.61	0.07
9 min.				
Wk. 1	9.54	11.18	—	1.65
	9.54	—	9.94	0.48
	—	10.91	9.71	1.56
Wk. 2	10.66	10.96	—	0.32
	10.66	—	9.75	1.04
	—	10.96	9.75	1.40
13 min.				
Wk. 1	13.16	15.46	—	1.21
	13.16	—	12.70	0.51
	—	15.39	12.66	2.09
Wk. 2	14.84	14.09	—	0.81
	15.37	—	14.06	1.29
	—	14.29	14.06	0.15

* $p < .05$.

TABLE 2

MEANS AND *t* VALUES (Paired Comparisons) FOR PRODUCTION ESTIMATES BETWEEN PULSE RATE CATEGORIES

	Beats/Min			*t*
	≤99	100–119	≥120	
4 min.				
Wk. 1	4.37	4.58	—	0.41
	4.10	—	4.78	0.93
	—	4.96	5.26	0.70
Wk. 2	5.83	4.93	—	1.36
	6.08	—	6.13	007
	—	4.90	6.13	2.81*
9 min.				
Wk. 1	8.82	10.03	—	2.08
	4.90	—	9.65	5.09**
	—	10.38	9.88	0.60
Wk. 2	10.64	10.78	—	0.01
	10.10	—	10.70	0.63
	—	11.10	10.70	0.26
13 min.				
Wk. 1	13.92	12.18	—	2.14
	13.55	—	14.40	0.37
	—	12.00	12.07	0.03
Wk. 2	16.13	14.94	—	0.96
	15.43	—	14.32	0.06
	—	14.70	15.58	0.16

* $p < .05$.
** $p < .01$.

DISCUSSION

The present experiments confirm the findings of earlier investigations that there is little evidence for an associated change in time estimation with increased pulse rate.

Taking the results of Francois (1927) and Hoagland (1933) as a guide to the prediction of the direction of change in time judgments with increases in body temperature, it was expected that, in the experiments here reported, with raised body temperature there would be a *decrease* in the time taken to count to 45 or 100 at an estimated rate of 1 digit per second; a *decrease* in the time estimated by the production method to be equivalent to 4, 9, or 13 min.; and an *increase* in the magnitude of the verbal estimates of the time felt to have passed during 4, 9, or 13 min. In none of the three experiments were these predictions statistically supported.

Several differences in experimental variables may account for these discrepancies. In each of the earlier experiments, without any detailed statistical examination of the data, conclusions were based on the responses of only three *S*s. In the present experiments a larger number of *S*s was examined and at normal and at raised body temperatures statistically significant inter-*S* variability was found in Exp. II and III. In addition to this variability, in all the present experiments individuals differed in the extent *and the direction* of change of estimate size with increases in body temperature.

Francois (1927) having found marked day to day variations in estimate size within each *S*, confined his comparisons within periods of approximately 30–40 min. Hoagland (1933) noted the time estimation of his two influenzal *S*s over a period of 48 hr., or less. In the present experiments therefore, any tendency for estimate size to vary with increased oral temperature would have needed to be quite marked to show above fluctuations in judgments from day to day. This did not, in fact, occur.

Both Francois (1927) and Hoagland (1933) studied the effect of raised body temperatures on the estimation of relatively brief intervals, during which *S*s were engaged in a rhythmic activity (i.e., tapping or counting at an estimated rate). In the present Exp. II and III in the estimation of much longer time intervals *S*s were required to make a single judgment at the end of the estimation which could have correction factors incorporated in it based upon feelings of whether they were tending to under- or overestimate the interval. Unfortunately no data on subjective impressions or accounts of corrections were requested of *S*s in the present Exp. II and III.

Finally, a variable which may be of significance is the way in which increases in body temperature are induced. In the present experiments, body temperature was raised as a result of exposure to uncomfortably high temperature environmental conditions.

Despite these differences in experimental technique, however, the fact that in three experiments on a larger number of *S*s, who were in normal health, no clear association was found, must cast doubt on the universality of an association between time estimation and body temperature changes. The precise conditions under which such an association might be found need to be clarified, and an exact description of the relationship must await further evidence.

REFERENCES

Broadbent, D. E. Some effects of noise on visual performance. *Quart. J. exp. Psychol.*, 1954, **6**, 1–5.

Clausen, J. An evaluation of experimental methods of time judgment. *J. exp. Psychol.*, 1950, **40**, 756–761.

Francois, M. Contribution à l'étude du sens du temps: La température interne comme facteur de variation de l'appréciation subjective des durées. *Année psychol.*, 1927, **28**, 186–204.

Hoagland, H. The physiological control of judgments of duration: Evidence for a chemical clock. *J. gen. Psychol.*, 1933, **9**, 267–287.

Hoagland, H. *Pacemakers in relation to aspects of behaviour.* New York: Macmillan, 1935.

Hoagland, H. Consciousness and the chemistry of time. In H. Abramson (Ed.), *Problems of consciousness.* New York: Corlies, Macy, 1951.

HOAGLAND, H., & PERKINS, C. T. Some temperature characteristics in man. *J. gen. Physiol.*, 1935, **18**, 399–408.

KASTENBAUM, R. Time and the nervous system: Some psychological notes. *Psychol. Rep.*, 1959, **5**, 281–288.

LEONARD, J. A. Five choice serial reaction apparatus. *Med. Res. Council Appl. Psychol. Res. Unit Rep.*, 1959, No. 326.

PIERON, H. *The sensations: Their functions, processes and mechanisms.* London: Muller, 1945.

PROVINS, K. A., HELLON, R. F., BELL, C. R., & HIORNS, R. W. Tolerance to heat of subjects engaged in sedentary work. *Ergonomics*, 1962, **5**, 93–97.

RASHKIS, H. A., & RASHKIS, S. R. Note on determinants of individual tempo. *Percept. mot. Skills*, 1962, **14**, 406.

SCHAEFER, V. G., & GILLILAND, A. R. The relation of time estimation to certain physiological changes. *J. exp. Psychol.*, 1938, **23**, 545–552.

WALLACE, M., & RABIN, A. I. Temporal experience. *Psychol. Bull.*, 1960, **57**, 213–236.

WOODROW, H. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* London: Chapman & Hall, 1951.

(Received December 24, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 580–587

EFFECTS OF CONTROLLED CODING CUES IN SHORT-TERM MEMORY[1]

RICHARD H. LINDLEY

Trinity University

3 experiments investigated the effects of the presence or absence of recoding and decoding cues on short-term memory (STM) for 3-letter items. The retention intervals were 3, 13, 39 sec. in Experiments I and II; 3, 12, 36 sec. in Experiment III. The Ss were 36 men and women in Experiment I, 24 in Experiment II, 30 in Experiment III. The results were (a) that recoding cues facilitated STM with low meaningfulness but interfered with STM with high meaningfulness ($p < .01$), (b) decoding cues facilitated STM with low meaningfulness ($p < .01$).

Miller (1956) suggested that the length of the immediate memory span was a function of the number of units as organized by *S* (called "chunks") in the to-be-recalled message rather than of the amount of information in the message as objectively measured in bits. He described an experiment by Smith (Miller, 1956, pp. 93–95) on a single *S* in which the length of the memory span for binary digits (e.g., 0100110110 . . .) was increased from about 10 to about 40 digits through training; e.g., the training consisted of teaching *S* a code name for each possible quadruplet of binary digits (e.g., 0000 = 0, 0001 = 1, 0010 = 2). When the binary digits were presented in the memory span experiment, *S* recoded the digits according to their previously-learned code names; at recall *S* decoded the code names back into the correct sequence of binary digits.

A study by Murdock (1961) showed that the amounts retained for the short-term memory (STM) of items that are well below the immediate memory span may also be a function of the number of "chunks" of information in the to-be-recalled message. Using the technique for studying STM developed by Peterson and Peterson (1959), Murdock compared the retention of three-letter consonant syllables with the retention of either one or three common words. The results showed "little difference between the retention of three consonants and three words but marked differences between the retention of three 'chunks' and one word [Murdock, 1961, p. 620]."

One purpose of the present studies was to investigate STM for three-letter items as a function of cues designed to control the coding performed by *S*. It was hypothesized that retention in STM could be facilitated by presenting cues along with the to-be-recalled item which would enable *S* to recode the item into a larger "chunk" of information. A second purpose was to explore the effects of presenting cues at recall designed to control the decoding of the recoded "chunk" into the to-be-recalled item.

---

[1] Experiments I and II were conducted at the University of Michigan in the summer of 1962 while the author was a participant in the National Science Foundation Research Participation Program. The author wishes to thank the advisor for this research, Arthur W. Melton, for his encouragement and many helpful suggestions; and Roy Hardman for collecting the data for Exp. III.

EXPERIMENT I

Method

Experimental design.—A 3 × 3 × 2 factorial design was used; there were three retention intervals (approximately 3, 13, and 39 sec.), three levels of meaningfulness (Low, Medium, High), and recoding cues were either present or absent. Each *S* was tested with two different items in each of the 18 conditions; thus each *S* was tested with all of the 36 items used in this study. A given item was used in only one condition for an individual *S*. The items were counterbalanced over *S*s so that each item within a given level of meaningfulness appeared equally often in each of the 6 possible conditions for that item (i.e., 3-, 13-, and 39-sec. retention intervals; recoding cues present or absent).

Materials.—There were 12 items at each level of meaningfulness; the items were selected from Exp. I of Johnson (1962). The items either were words (High Meaningfulness) or were syllables (Medium Meaningfulness) or nonsyllables (Low Meaningfulness) as determined by *S*s' ratings (Johnson, 1962). Data on paired-associate learning where these items were response terms are available from Underwood and Schulz (1960).

All of the to-be-recalled three-letter items were either English words or parts of English words. The three letters to be recalled were typed in capital letters and the recoding cues consisted of additional lower-case letters in the appropriate positions to make a complete English word. For example, two items were CQU and WAS; the items with recoding cues present were acQUire and WAsh, respectively. The items used in Exp. I are presented in Table 1 with the recoding cues that were used; of course half the items were presented without recoding cues for each *S*. Table 1 also shows the total proportion correctly recalled for each item for the combined three retention intervals as a function of the presence or absence of recoding cues.

Two different forms of each of three basic lists were used; the two forms differed only in that all items with recoding cues present in one form had recoding cues absent in the other form and vice versa. The retention intervals were presented in the order 3, 13, and 39 sec. or some permutation of this order. The six conditions of meaningfulness and recoding cues present or absent were presented in a nonsystematic arrangement in each of the basic lists.

Procedure.—The items were presented on a memory drum at about a 1.6-sec. rate of presentation. The word "Ready" appeared and was followed by a blank space which in turn was followed by the to-be-recalled item. The *S* had been instructed to spell only the three capital letters both at the time the item was presented and at recall; thus the recoding cues were not spelled or pronounced aloud during the experiment. Following the item, a row of seven random digits appeared and *S* read these digits aloud one by one under

TABLE 1

ITEMS USED AND PROPORTIONS CORRECTLY RECALLED FOR EACH ITEM IN EXP. I

Low Meaningfulness			Medium Meaningfulness			High Meaningfulness		
Items	Recoding Cues		Items	Recoding Cues		Items	Recoding Cues	
	Absent	Present		Absent	Present		Absent	Present
eXPOse	.78	.89	CHIp	.94	.83	hITS	.94	.89
COMPASS	.78	.83	SING	1.00	1.00	WASh	.89	.83
hoTLY	.61	.78	SOUth	.89	1.00	CUTS	1.00	.78
nATION	.67	.72	UNDer	.94	.94	BUTter	.89	.94
aRCH	.67	.83	JOKe	.94	.89	MANy	1.00	.94
hOUSe	.94	.94	ROPe	.89	.83	WHOm	.94	.89
mOULD	.83	.72	STIck	.83	.89	HERe	1.00	.89
ALIke	.94	.83	ENTer	.78	.83	tURN	.94	.94
ABLe	.78	.72	JUMp	.89	.94	ARTist	1.00	.83
eMPTY	.78	.83	EIGht	.72	.72	SLED	.94	.78
rOMP	.67	.94	VERy	.72	.72	HATe	1.00	.94
acQUire	.50	.61	FROm	.94	.94	STYle	.83	.78

Note.—All items are shown with recoding cues present.

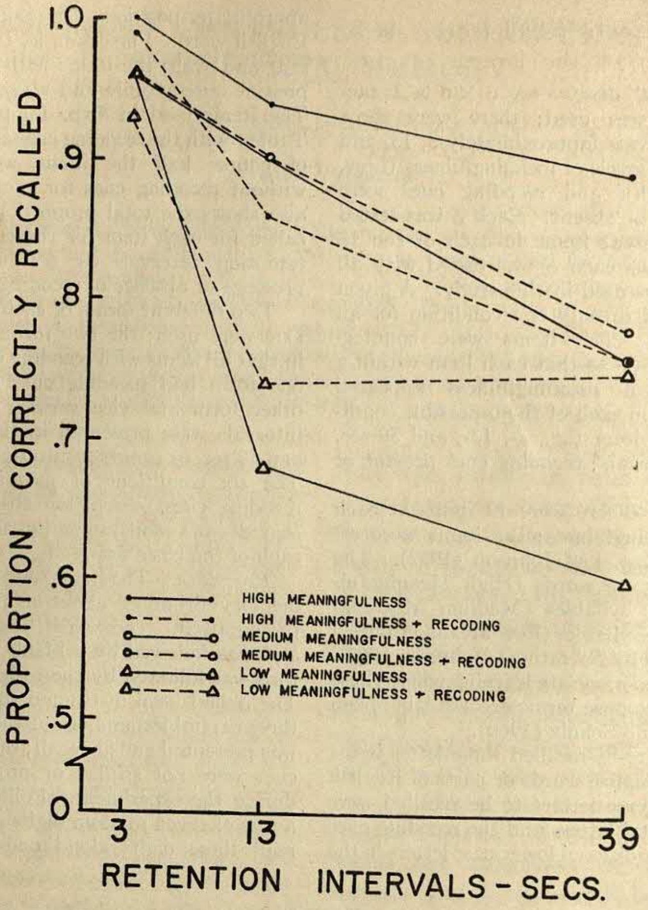


FIG. 1. Proportions correctly recalled in Exp. I.

instructions to perform this task as fast as he could and not to rehearse the item during this time. Subsequent rows of seven random digits were read under the same instruction for the appropriate length of time until the cue for recall (three asterisks) appeared. Approximately 13 sec. were allowed for *S* to spell the item during the recall interval. Credit was given for a correct response if *S* corrected an error during the recall interval. After the recall interval, the word "Rest" appeared and a 6-sec. rest was given.

Six *S*s were tested with each of the two forms of the three basic lists. Different *S*s were started at three different places in each form of the basic list. The *S*s were given two rest periods equally spaced apart during the experiment of about 1½ min. duration while the tape was changed.

Subjects.—Thirty-six paid volunteers from the University of Michigan served as *S*s. Immediately preceding this experiment all *S*s had participated in an STM study using the original Peterson and Peterson (1959) technique except that the three-consonant items of low meaningfulness were presented visually. Approximately 5 min. elapsed between the two studies during which time the memory drum tapes were changed and the instructions for this study were read. The first 12 *S*s were randomly assigned to the two forms of the first basic list; the next 12 *S*s were randomly assigned to the two forms of the second basic list, etc.

Results and Discussion

Figure 1 presents the data on the proportions of items recalled correctly. Inspection of Fig. 1 shows that the proportion correctly recalled decreased as a function of the length of the retention interval and as a function of decreasing meaningfulness, *F*s (2, 595) = 44.43 and 20.19, $p < .01$, respectively. The presence or absence of recoding cues was not a significant source of variance by itself, *F* (1, 595) < 1; however, the Meaningfulness × Recoding Cues interaction was significant, *F* (2, 595) = 5.05, $p < .01$. As Fig. 1 shows, the presence of recoding cues facilitated STM for the low meaningfulness items, interfered with retention of the high meaningfulness items, and had no effect on STM for the medium meaningfulness items. The only other significant difference was Retention Intervals × Meaningfulness, *F* (4, 595) = 3.43, $p < .01$; this represents a difference in slope of the retention curves as a function of meaningfulness which can be seen in Fig. 1.

Twenty of the items used in this study were also response terms in a paired-associate task of Underwood and Schulz (1960, Exp. 12). The product-moment correlation between the number of correct anticipations in 15 trials in that study and the proportion correctly recalled with recoding cues absent in this study was 0.69, $df = 18$, $p < .01$. Thus degree of meaningfulness has a similar effect in STM and in paired-associate learning; this finding is in complete accord with the findings of Peterson, Peterson, and Miller (1961).

The differential effect of recoding cues on STM for high and low meaningfulness items may be explained by the ad hoc postulation of two opposed factors which jointly help to determine recall. The first is a recoding or "chunking" factor; it is assumed that the recoding cues permit *S* to recode the items into larger "chunks" of information. The second and opposed factor is interference; it is assumed that the extra letters of the recoding cues compete with the to-be-recalled three letters at recall (and perhaps during the original presentation of the item as well) which leads to poorer retention. For the high meaningfulness to-be-recalled three-letter words, the interference factor is assumed to be dominant; the recoding factor is not useful because the three-letter words are already coded as one chunk. The interference effects are also assumed to be present for the low meaningfulness items but such effects are assumed to be overriden by the beneficial chunking effects of the recoding cues. While this analysis of the effects of recoding cues is admittedly after the fact, it does lead to further research.

One implication of the above analysis is that cues given at recall which permit *S* to discriminate between the to-be-recalled three-letter item and the recoding cues should facilitate STM by reducing response competition from the item and the cues. This was tested in Exp. II.

EXPERIMENT II

Method

Experimental design.—A 3 × 2 × 2 factorial design was used; there were three retention intervals (approximately 3, 13, and 39 sec.), recoding cues were either present or absent, and decoding cues were either present or absent. As in Exp. I each *S* was tested with two items in each condition; a given item was used in only one condition for an individual *S*; and the items were counterbalanced over *S*s so that each item appeared equally often in each of the 12 experimental conditions.

Materials.—There were 24 three-letter items of low meaningfulness; all items appeared in the middle of words which occurred 20 times per million or more in English text as estimated by the G (General) count of Thorndike and Lorge (1944). With two exceptions all items were composed of three consonant letters. Furthermore, each item "cut across" a syllable; i.e., the items were not pronounced as a whole in standard English pronunciation.

As in Exp. I, the recoding cues were lowercase letters which, in conjunction with the to-be-recalled three capital letters, formed a word. The decoding cues consisted of

TABLE 2

ITEMS USED IN EXP. II AND III

Items	
1. douBTLess	16. eXCHange
2. el[illegible]TRic	17. maRBLe
3. ju[illegible]Ment	18. aNSWer
4. hi[illegible]Way	19. buTCHer
5. chil[illegible]en	20. moSTLy
6. faRTHer	21. neiGHBor
7. asseMBLy	22. newSPaper
8. desCRibe	23. ORCHard
9. frieNDLy	24. OXYGen
10. crYSTal	25. ciRCLe
11. eXPLain	26. saNDWich
12. ligHTNing	27. reCKLess
13. baCKWard	28. puRPLe
14. eMPTy	29. aLTHough
15. ceNTRal	30. iMPROve

Note.—All items are shown with recoding cues present.

The fi[illegible] items were used in Exp. II. All 30 items w[illegible] Exp. III.

presenting the decoding cue lower-case letters at recall instead of the three asterisks. For example, one of the to-be-recalled items was BTL; with recoding cues present *S* saw douBTLess; when decoding cues were presented during the recall interval, *S* saw dou——ess. The items used in Exp. II are presented in Table 2 with the recoding cues used; the decoding cues were simply the lower-case letters with the capitalized letters removed.

Two different forms of each of six basic lists were used; the two forms of each list bore the same relationship to each other as occurred in Exp. I. The retention intervals and the conditions with recoding and decoding cues present or absent were arranged nonsystematically in a manner similar to Exp. I.

Procedure.—The procedure of Exp. I was followed [illegible] closely. Two *S*s were tested with each of the two forms of the six basic lists and two different starting places were used for each form. The decoding cues were presented only during the first 5 sec. of the 13-sec. recall interval.

Subjects.—Twenty-four paid volunteers from the University of Michigan served as *S*s. The only difference between this study and Exp. I with respect to *S*s was that in Exp. II the first four *S*s were randomly assigned to one of the two forms of the first basic list, etc.

Results and Discussion

Figure 2 presents the data on the proportion of items recalled correctly; once again there was a significant amount of forgetting over retention intervals, $F\ (2, 253) = 20.00, p < .01$. As Fig. 2 shows, the presence of recoding cues with items of low meaningfulness significantly facilitates recall in STM, $F\ (1, 253) = 36.50$, $p < .01$; thus the findings of Exp. I for the low meaningfulness items are replicated. The presence of recoding cues also affected the slope of the retention curves as is shown by the significant Retention Intervals × Recoding Cues interaction, $F\ (2, 253) = 4.13, p < .05$. Furthermore recall of low meaningfulness items in STM is facilitated by the presence of decoding cues, $F\ (1, 253) = 12.53$, $p < .01$; however, it seems likely that two factors are involved in such facilitation; viz., reduction in interference and guessing.

It has been assumed that the presence of decoding cues at recall which have been preceded by recoding cues leads to better recall by permitting *S* to discriminate between the to-be-recalled item and the recoding cue letters and to reject the latter. The results of the condition with both recoding and decoding cues present support this interpretation. However, *S* may be able to guess the correct item when decoding cues are present without having first seen the item with recoding cues present. As a control for this possibility, the condition with the three-letter items followed by decoding cues at recall was included in the design of Exp. II. It has been suggested to the author that such a control may not be adequate because the three-letter item response may interfere with the responses elicited by the decoding cues at recall and hence this control may underestimate the proportion of items that can be "recalled" (by guessing) when decoding cues are available. The control for this possibility is to include a condition where no item is presented first and *S* is presented with decoding cues at recall under instructions to fill in those

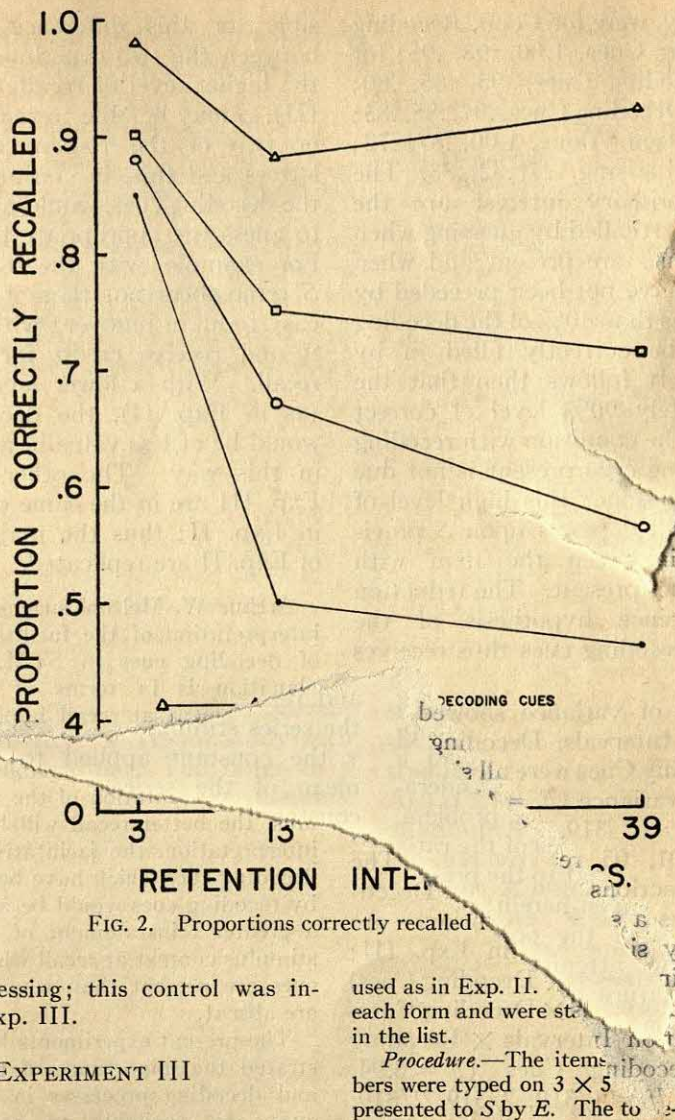


FIG. 2. Proportions correctly recalled

cues by guessing; this control was included in Exp. III.

EXPERIMENT III

Method

Experimental design and materials.—Experiment III used the same 3 × 2 × 2 factorial design with the same factors as were used in Exp. II. In Exp. III the retention intervals were 3, 12, and 36 sec. and 6 sec. were allowed for recall. The *S* was instructed to fill in the decoding cues by guessing. The 30 items used in Exp. III are presented in Table 2; a given item was used in only one condition for an individual *S* and the items were counterbalanced over *S*s as in Exp. I and II. The same procedure with respect to basic lists, forms of basic lists, and the arrangement of experimental conditions was used as in Exp. II. ted with each form and were st ne place in the list.

Procedure.—The items dom numbers were typed on 3 × 5 ds and were presented to *S* by *E*. The to -recalled item was presented for approximately 1.5 sec. The various time intervals were controlled by *E* by observing a Standard Electric timer. No long rests were given *S* during the experiment.

Subjects.—The *S*s were 30 volunteer students from Trinity University. The *S*s had not previously participated in an STM study and were assigned to lists as in Exp. II.

Results and Discussion

The proportions correctly recalled for the 3-, 12-, and 36-sec. intervals, respectively, were for Cond. Recoding + Decoding Cues, 1.00, .98, .95; for Cond. Recoding Cues, .95, .85, .80; for Cond. Decoding Cues, .97, .85, .85; for Cond. Item Alone, 1.00, .87, .72; for Cond. Guessing, .27, .32, .38. The data of primary interest are the proportions recalled by guessing when decoding cues are present and when such cues have not been preceded by an item; less than 40% of the decoding cues can be correctly filled in by guessing. It follows then that the approximately 90% level of correct recalls for the condition with recoding and decoding cues present is not due to guessing alone; this high level of correct recall depends upon *S* previously having seen the item with recoding cues present. The reduction of interference hypothesis of the action of decoding cues thus receives support.

Analysis of variance showed[illegible] pre- saw Retention Intervals, Decoding[illegible] II are and Recoding Cues were all [illegible]oding cues sources of variance (F = [illegible] simply the 5.00, $df = 2/319$, [illegible]apitalized letters $p < .01$, .01, .05, re of each of six The three interactions [illegible] forms of each Decoding Cues as a [illegible]ship to ea[illegible]ariance are statistically si[illegible] I. The ret[illegible] [illegible]s with reco[illegible] in Exp. III; i.e., Recodi[illegible]or absent × Decoding Cues, Retentio[illegible] in a m[illegible]s × Decoding Cues, and Reten[illegible] Intervals × Recoding Cues × Decoding Cues ($F = 3.94$, 3.38, 8.31, $df = 1/319$, 2/319, 2/319, $p < .05$, .05, .01, respectively). None of the latter three interactions was significant in Exp. II. A comparison of the retention curves in the two experiments for the condition where the three-letter item was presented first and was followed by decoding cues at recall shows that recall for this condition was higher relative to the other groups in Exp. III than in Exp. II. The higher level of recall in Exp. III for this group is apparently responsible for this difference in results between the two experiments. With the higher level of recall (as in Exp. III), *S* may be able to remember one or two of the to-be-recalled three letters and this, in conjunction with the decoding cues, would make it easy to guess the appropriate third letter. For example, with the item DGM, if *S* remembers DG, then it should be easy to fill in juDG–ent with the letter M and receive credit for a correct recall. With a lower level of recall (as in Exp. II), the decoding cues would be of less value in aiding recall in this way. The other results in Exp. III are in the same direction as in Exp. II; thus the major findings of Exp. II are replicated.

Arthur W. Melton[2] suggested another interpretation of the facilitative effects of decoding cues in STM. This ex-[illegible]lanation is in terms of a stimulus of dec[illegible]ment at recall hypothesis; i.e., been preceded by [illegible]recoding [illegible]situation to better recall by [illegible]able during the discriminate [illegible]tion of the verbal stim-[illegible], the better recall will be. By this interpretation, the facilitative effects of decoding cues which have been preceded by recoding cues would be attributed to a greater reinstatement of the original stimulus context at recall when decoding cues are present than when these cues are absent.

The present experiments have demonstrated the importance of the recoding and decoding processes in STM when cues are available to control these processes. Further, the effects seem to be a function of the level of meaningfulness of the items to be recalled. Miller's (1956) suggestion that "it seems probable that even memorization can be studied in these terms (i.e., recoding) [p. 95]" appears to be essentially correct. The present experiments have investigated the recoding and decoding processes only under controlled coding conditions;

---

[2] A W. Melton, personal communication, 1962.

further research is needed when recoding and decoding cues are not available to S and he provides his own.

REFERENCES

JOHNSON, R. C. Reanalysis of "Meaningfulness and verbal learning." *Psychol. Rev.*, 1962, **69**, 233–238.

MILLER, G. A. The magical number seven; plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

MURDOCK, B. B., JR. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

PETERSON, L. R., PETERSON, M. J., & MILLER, A. Short-term retention and meaningfulness. *Canad. J. Psychol.*, 1961, **15**, 143–147.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia University, Bureau of Publications, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received December 29, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 588–595

NEW DIMENSIONS OF THE INTERMEDIATE SIZE PROBLEM: NEITHER ABSOLUTE NOR RELATIONAL RESPONSE[1]

MICHAEL D. ZEILER[2]

New School for Social Research

In certain specific instances the ratio theory predicted that response would be to a previously negative stimulus although the positive intermediate training stimulus was the intermediate member of the test set. Experiment I tested this prediction (N = 168) with young children and revealed significant preference for the predicted stimulus. Experiment II (N = 60) tested 2 combinations in an attempt to replicate the effect found in Experiment I; however, prediction was not supported. Experiment III (N = 120) did replicate the phenomenon and also supported the contention of the ratio theory that with training held constant with 1–2–8 there would be increasing preference for Stimulus *1* as the test set was shifted from 1–2–6 to 1–2–5 to 1–2–4 to 1–2–3. The direction of modification of the ratio theory to account for the negative data was indicated. Abstraction, as indexed by verbalization, appeared to be irrelevant to choice performance.

The ratio theory of intermediate size discrimination, a formulation based essentially on the underlying assumptions of Helson's (1959) adaptation-level theory, has provided a conceptual framework for consideration of the intermediate size problem. Only the main hypotheses of the ratio theory that are relevant to the present experiments are given herein; a complete discussion of the position, its relation to other attempts to define the essential nature of the stimulus in transposition studies, and the deduction of the results of the majority of experiments on the intermediate size problem has been presented elsewhere (Zeiler, 1963). The hypotheses relevant to the current series of investigations are as follows:

1. The adaptation level (AL) is the weighted log mean of series stimuli and residual AL. Let X_i represent the series stimuli, R the residual AL, y the constant applied to the log mean of the series stimuli, x the constant applied to the residual AL, $x + y = 1.00$, and n the number of series stimuli. The AL formula is:

$$\text{Log AL} = y\left(\frac{\Sigma \log X_i}{n}\right) + x \log R$$

(*a*) The S enters training with a residual AL which exerts an extra-experimental effect upon the series AL. With each successive trial, R becomes increasingly a function of the series stimuli since R is primarily dependent upon immediate past stimulation. The effect of irrelevant prior stimulation is negligible by the completion of training so that:

$$\text{Log training AL} = \frac{\Sigma \log X_i}{n}$$

(*b*) The AL on the first test trial after the completion of training is expressed by the formula:

---

[1] This paper is based on part of a dissertation submitted to the Graduate Faculty of the New School for Social Research in partial fulfillment of the requirements for the PhD degree. The writer wishes to express his appreciation to the chairman of his dissertation committee, Jerome Wodinsky, for his advice and interest during all phases of this study.

[2] Now at Wellesley College.

$$\text{Log test AL} = y\left(\frac{\Sigma \log X_i}{n}\right) + x \log \text{training AL}$$

2. Each stimulus in the series is defined as the ratio of the stimulus area to the AL.

3. The *S* learns in training to respond to the ratio of the positive stimulus to the AL.

4. Probability of response on the first test trial is a function of the degree of similarity between the individual test stimulus ratios and the positive training ratio.

Zeiler (1963) suggested that the values of the constants for work with young children might be assumed to be .40 and .60 for x and y, respectively, based upon an examination of the range of these constants which would account for the Stevenson and Bitterman (1955) study with preverbal children. The range of y values (since $x + y = 1.00$, reference need only be made to y) adequate to deduce these results was .60–.80; the value of .60 was arbitrarily selected for the purpose of prediction of response with 4- and 5-yr.-old children.

If, after training to the intermediate member of a set of three stimuli, the test set includes the stimulus that was positive in training as well as one of the negative stimuli, absolute theory (Spence, 1942) predicts that the response on the first test trial could be to either the formerly positive or to the newly introduced element. Which of these two stimuli would be chosen depends upon the shape and extent of the hypothesized generalization gradients. In no event should the response be to the third stimulus which was present but not positively reinforced in training since the mechanism underlying original learning is the development of greater effective excitatory potential for the positive element in relation to either of the negative members. In the same situation, if the intermediate member in training continues to be the intermediate element of the test set, relational theory must predict continued response to that stimulus regardless of the absolute sizes of the stimuli in the sets. Only by post hoc assumptions of learned subrelationships between pairs of training set stimuli could any other response be explained.

In certain specific instances of the above condition, the ratio theory (assuming a y value of .60) makes a prediction that is contrary to that of any currently formulated absolute or relational theory. The predicted response is to the nonintermediate stimulus that was negative in training. For example, if training is with stimuli 4 in.^2, 5.6 in.^2, and 42.2 in.^2, the AL is 9.8 so that the positive training ratio (the ratio of 5.6 to the AL) is .57. With a test set composed of stimuli 4 in.^2, 5.6 in.^2, and 7.8 in.^2, the first test trial AL is 7.01 so that the test stimuli have the following ratio values: 4 $\text{in.}^2 = .57$, 5.6 $\text{in.}^2 = .80$, 7.8 $\text{in.}^2 = 1.11$. Since .57 is the test ratio closest to that of the positive training ratio, predicted response on the first test trial is to the 4-in.^2 stimulus.

EXPERIMENT I

Method

Materials.—The stimuli were eight unpainted squares cut from ¼-in. Masonite. An indented area was cut in the bottom of each block so that the goal object, a small red plastic chip, could be placed under the block without being evident. The blocks were designated in order of size with the numbers 1–8: Block 1 had a surface area of 4 in.^2 and the corresponding values for each successive block in the series increased by a factor of 1.4, the largest, "8," being 42.2 in^2.

An opaque screen, 16 × 20½ in., was interposed between *S* and the blocks while the blocks were being arranged for each trial. Both the screen and a tray, 20 × 30 in., on which the blocks were arranged were made of heavy buff-colored cardboard.

Procedure.—The spatial arrangement of the three blocks used in each trial was predetermined by use of a table of random numbers. The only limitation on randomness was that the intermediate-sized block could not appear in the same position on more than two successive trials. The *S* was seated before a table on which the tray was placed and was informed that *E* would hide the red toy under one of the blocks. The screen was then placed in front of the blocks, the predetermined spatial arrangement was established, and the chip was placed under the intermediate-size block. The blocks were arranged about 2 in. apart with their centers approximately lined up. The screen was then removed and *S* was instructed to find the goal object. On the first five trials *S* was allowed to look under each block until the chip was found (correction method). Commencing with the sixth trial, *S* was allowed a single choice (noncorrection method). Training was continued to a criterion of five consecutive first choices of the intermediate-sized block.

Immediately upon reaching criterion the test set blocks were substituted for those used in training and a chip was placed under each of the three elements. The first choice was recorded. Each *S* was used in one training and test combination and for a single test trial. At the conclusion of the test trial, *S*'s ability to abstract the concept of middle-size was investigated. Each child was asked to identify the middle-sized, the largest, and the smallest block. Questioning was continued until it became evident that *S* either did or did not know which was the intermediate-sized member of the series. Those that either verbalized the concept or could consistently identify the intermediate-sized block were classified as capable of abstraction.

Notation system.—Training and test combinations are described by first noting the stimulus numbers comprising the training set and then those of the test set: the sets are separated by a colon. For example, if *S* was given Stimuli 1, 2, and 8 in training and Stimuli 1, 2, 3 in test, this is shown as 1-2-8:1-2-3.

Subjects.—The *S*s were 168 4- and 5-yr.-old children from day care centers in New York City. They were assigned to groups randomly with 60 *S*s in Group E and 36 *S*s in each of Groups CI, CII, and CIII. Three *S*s were run on each training and test combination.

Stimulus sets.—Predicted response in test was to the italicized element.

Group E. Training and test sets were such that the test set differed from the training set only in that the most extreme stimulus used in training was replaced in test.

1-2-6:*1*-2-3
1-2-7:*1*-2-3
1-2-7:*1*-2-4
1-2-8:*1*-2-3
1-2-8:*1*-2-4
1-2-8:*1*-2-5
2-3-7:*2*-3-4
2-3-8:*2*-3-4
2-3-8:*2*-3-5
3-4-8:*3*-4-5
1-5-6:4-5-*6*
1-6-7:4-6-*7*
1-6-7:5-6-*7*
1-7-8:4-7-*8*
1-7-8:5-7-*8*
1-7-8:6-7-*8*
2-6-7:5-6-*7*
2-7-8:5-7-*8*
2-7-8:6-7-*8*
3-7-8:6-7-*8*

Group CI. Test sets were identical to Group E and training set was varied so that predicted response was to the opposite extreme stimulus predicted for Group E.

1-4-6:1-2-*3*
1-5-7:1-2-*4*
1-6-8:1-2-*5*
2-5-7:2-3-*4*
2-6-8:2-3-*5*
3-6-8:3-4-*5*
1-3-6:*4*-5-6
1-3-7:*4*-6-7
1-3-8:*4*-7-8
2-4-7:*5*-6-7
2-4-8:*5*-7-8
3-5-8:*6*-7-8

Group CII. Test sets were identical to Groups E and CI and training set was varied so that predicted response was to the intermediate stimulus.

1-3-6:1-*2*-3
1-3-7:1-*2*-4
1-3-8:1-*2*-5
2-4-7:2-*3*-4
2-4-8:2-*3*-5
3-5-8:3-*4*-5
1-4-6:4-*5*-6
1-5-7:4-*6*-7
1-6-8:4-*7*-8
2-5-7:5-*6*-7
2-6-8:5-*7*-8
3-6-7:6-*7*-8

Group CIII. Training sets were identical to Group E while test sets differed from Group E in that the new stimulus introduced in test was further removed in size from the intermediate.

1-2-6:1-*2*-4
1-2-7:1-*2*-5
1-2-8:1-*2*-6
2-3-7:2-*3*-5
2-3-8:2-*3*-7
3-4-8:3-*4*-6
1-5-6:3-*5*-6
1-6-7:3-*6*-7
1-7-8:3-*7*-8
2-6-7:4-*6*-7
2-7-8:4-*7*-8
3-7-8:5-*7*-8

Results

There was no significant difference in number of trials to criterion for the four groups, F (3, 164) = 1.84. The principal results are presented in two ways: (*a*) Number of *S*s within each group responding as predicted and number of *S*s responding otherwise; (*b*) Number of training-test combinations in which at least two of the three *S*s responded as predicted and the number of combinations in which majority response did not support prediction.

The results are given in Table 1 for each scoring method. Of the 168 *S*s in the four groups, 132 responded as predicted, χ^2 (1) = 54.84, $p < .001$. When analyzed in terms of *S*s, response was significantly as predicted in Group E, χ^2 (1) = 17.06, $p < .001$; Group CI, χ^2 (1) = 28.4, $p < .001$; and Group CIII, χ^2 (1) = 16.0, $p < .001$. The only group in which response was not significantly in accord with prediction was Group CII: χ^2 (1) = 1.78, $.20 > p > .10$. It should be noted that the χ^2 test was based on an expected frequency of 50% of the cases supporting and 50% opposing prediction with the latter classification including response to either of the nonpredicted test stimuli. This is an extremely stringent requirement and may not do justice to the accuracy of the predictions. If the expected frequency were to be 33% responding as predicted and 67% responding opposed to prediction, the data for Group CII were significant, χ^2 (1) = 12.5, $p < .001$.

When analyzed in terms of number of training-test combinations in which at least two of the three *S*s responded as predicted, Groups E, CI, and CIII showed a significant number of conditions in which majority response was to the predicted stimulus. Group E, χ^2 (1) = 9.8, $p < .01$; Group CI, χ^2 (1) = 12.0, $p < .001$; Group CIII, χ^2 (1) = 8.34, $p < .01$. In Group CII, the frequency of majority response was not significant, χ^2 (1) = 3.00, $.10 > p > .05$. The combinations for Group E yielding majority response not as predicted were 1-5-6:4-5-6, 2-6-7:5-6-7, and 1-2-8: 1-2-5. For both scoring methods, the data of the three control groups reveal that nothing peculiar to either the particular training or test sets was responsible for the results in Group E.

Thirty-three *S*s were capable of abstracting the concept of intermediate size. Of these 33, 27 responded as predicted, χ^2 (1) = 13.2, $p < .001$. Of the 27 responding as predicted, 8 *S*s chose the large stimulus in test, 5 *S*s chose the intermediate stimulus, and 14 *S*s chose the small. Five of the 6 *S*s capable of abstraction and failing to respond as predicted showed intermediate response while 1 chose the small stimulus.

TABLE 1

FIRST TEST TRIAL RESPONSE IN TERMS OF PREDICTION

Response	Group E		Group CI		Group CII		Group CIII	
	Ss	Sets	*Ss*	Sets	*Ss*	Sets	*Ss*	Sets
As predicted	46	17	34	12	22	9	30	11
Not as predicted	14	3	2	0	14	3	6	1

TABLE 2

RESULTS

Training Set	Test Set	Trials to Criterion		Frequency of First Test Trial Response		
		M	*SD*	S	I	L
1-2-8	1-2-5	13.1	3.51	8	21	1
1-7-8	6-7-8	13.4	4.18	5	13	12

EXPERIMENT II

Experiment I supported the prediction of the ratio theory that in certain cases test response would be to a formerly negative stimulus even though the formerly positive stimulus was present and continued to be the intermediate member of the test set. Since only three *S*s per condition were employed in Exp. I, the importance of the phenomenon revealed by that study made it advisable to select a small sample of the 20 combinations given to Group E in that investigation and test them with a relatively large population.

Method

The general procedure was identical to that of Exp. I. The *S*s, 60 4- and 5-yr.-old children from day care centers in New York City, were assigned randomly to two conditions with 30 *S*s in each.

One condition was selected randomly from the three that yielded response contrary to prediction in Group E of Exp. I and one was selected from the 17 sets that supported prediction. The condition selected from the negative cases was 1-2-8:1-2-5 and that chosen from the supporting sets was 1-7-8:6-7-8. In the first case predicted response was to Stimulus *1* while in the second situation predicted response was to Stimulus *8*.

Results

Mean trials to criterion, standard deviation of trials to criterion, and the results for both combinations are given in Table 2. In neither case was the predicted stimulus chosen by the majority of *S*s. 1-2-8:1-2-5 showed that response was significantly more frequent to the nonpredicted stimuli, χ^2 (1) = 6.52, $p < .02$. Considering the three test stimuli independent of classification by prediction, there was a significant difference (expected frequency per cell = 10) in number of responses to each, χ^2 (2) = 20.6, $p < .001$. Evaluation of number of responses to the predicted stimulus (*1*) against the stimulus chosen most often (*2*) revealed significant preference for Stimulus *2*, χ^2 (1) = 5.82, $p < .02$. With 1-7-8:6-7-8 there were no significant differences either in terms of predicted and nonpredicted response, consideration of each stimulus separately, or by comparison of the predicted test stimulus (*8*) with the most frequently chosen stimulus (*7*). While 1-2-8:1-2-5 replicated the findings of the first experiment in showing response contrary to theoretical expectation, 1-7-8:6-7-8 not only failed to support prediction but the apparent equal response to Stimuli *7* and *8* failed to replicate the findings for this combination in Exp. I where majority response had been to Stimulus *8*.

Twelve *S*s were capable of abstracting the concept of intermediate size. Of these 12, 7 responded to the intermediate test stimulus.

EXPERIMENT III

The results of Exp. II made it necessary to investigate further the conditions under which the ratio theory can account for the nature of the response in the intermediate size problem with the values of the *x* and *y* constants in the adaptation-level formula assumed to be .40 and .60, respectively. Data of a systematic nature were required to determine the precise point at which these constant values resulted in erroneous prediction. Furthermore, the original purpose of Exp. II was not yet fulfilled: to demonstrate with a large population the phenomenon revealed in Exp. I; that is, choice of a previously negative training stimulus although the previously positive intermediate stimulus was present in test and continued to have the relationship of intermediate size to the other two stimuli with which it appeared.

Clearly negative data were yielded by 1-2-8:1-2-5 where response was predicted to Stimulus *1* but in fact occurred to *2* in both Exp. I and II. This case was therefore taken as the reference point for the current study: training was held constant with 1-2-8 with test on either 1-2-3, 1-2-4, 1-2-5, or 1-2-6. Table 3 indicates the test stimulus ratios and predicted first test trial response for the four combinations (positive training ratio = .57). This table reveals that response was predicted to Stimulus *1* with Test Sets 1-2-3, 1-2-4, and 1-2-5 and to Stimulus *2* with Test Set 1–2–6. Such a study had to yield information as to the precise point of breakdown of the predictions of the ratio theory: it was already known that the theory was inadequate to explain 1-2-8:1-2-5.

Another basic prediction of the ratio theory is that as test is shifted from 1-2-6 to 1-2-5 to 1-2-4 to 1-2-3 there should be increasing frequency of response to Stimulus *1* as that stimulus ratio progressively approaches the positive training ratio relative to the other test stimulus ratios. Evidence for an absolute preference for Stimulus *1* in Test 1-2-3, 1-2-4, or 1-2-5 as well as an orderly shift in frequency of response to that stimulus as the test set was changed progressively from 1-2-6 to 1-2-3 would be of considerable importance since it is not readily apparent as to how either absolute or relational theory could deduce either phenomenon.

Method

The general procedure was identical to that of Exp. I. The *S*s, 120 4- and 5-yr.-old children from day care centers in New York City, were assigned randomly to four groups with 30 *S*s in each. All *S*s were trained with 1-2-8. There were four different test sets administered as follows. Group 1: 1-2-3; Group 2: 1-2-4; Group 3: 1-2-5; Group 4: 1-2-6.

Results

There was no significant difference in number of training trials to criterion for the four groups, F (3, 116)

TABLE 3

PREDICTIONS AND RESULTS

Group	Test Set	Test AL	Test Stimulus Ratios			First Test Trial Response		
			S	I	L	S	I	L
1	1-2-3	7.01	.57[a]	.80	1.11	16	10	4
2	1-2-4	7.5	.53[a]	.75	1.47	10	18	2
3	1-2-5	8.04	.50[a]	.70	1.92	9	21	0
4	1-2-6	8.57	.47	.65[a]	2.51	3	24	3

Note.—Positive training ratio = .57.

[a] Ratio of predicted stimulus for each group.

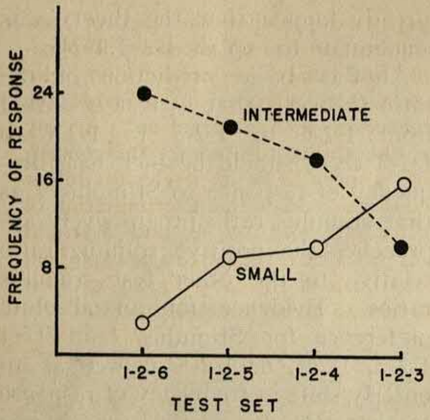


FIG. 1. Distribution of responses on first test trial to the small and the intermediate test stimulus.

= .37. The number of responses to each stimulus in test is given in Table 3, overall χ^2 (6) = 18.67, $p < .01$. There was a significant difference in number of responses to each stimulus for each group (expected frequency per cell = 10), Group 1: χ^2 (2) = 7.2, $p < .05$; Group 2: χ^2 (2) = 12.8, $p < .01$; Group 3: χ^2 (2) = 22.2, $p < .001$; Group 4: χ^2 (2) = 30.4, $p < .001$. Groups 1 and 4 supported prediction while Groups 2 and 3 did not. The distribution of small and intermediate responses for each group is indicated graphically in Fig. 1. The data reveal replication of the major effect of Exp. I (as shown by Group 1) as well as confirmation of the prediction of increased frequency of response to Stimulus *1* with progressive alteration of test sets.

Twenty-six *S*s were able to abstract the concept of middle-size. Of these 26, 12 responded to the intermediate stimulus in test.

DISCUSSION

Group 1 of Exp. III replicated with a larger *N* the phenomenon found in Exp. I. A new dimension has been found in the intermediate size problem, namely, systematic response to a test stimulus that is neither the absolute stimulus of the training set nor that which has the same relationship to the test set as had the positive training stimulus to the other elements with which it appeared. This aspect of the intermediate size problem, predicted by the ratio theory, would appear to be inexplicable in terms of either relational or absolute conceptions.

While Groups 1 and 4 of Exp. III supported the predictions of the ratio theory, Groups 2 and 3 as well as both conditions of Exp. II did not. An initial possibility for reconciling discrepant data is a reconsideration of the constants employed in the adaptation-level formula. All of the data of Exp. III (positive as well as negative) and II can be deduced by changing the value of the y constant from .60 to .30 ($x = 1 - y$). However, in order to explain the results of Stevenson and Bitterman (1955) where children of the same age level were employed, as well as to explain the results of Exp. I, it is not possible to assume that these are the general constant values. The possibility remains that the constants are not fixed values for a given species but are a function of specific stimulus set variables. The most obvious variable in the current experiments is stimulus set symmetry: the relationship of the intermediate to the two extreme stimuli. Perfect symmetricality would be the case in which the intermediate stimulus is equally different in area from the two extremes while asymmetry would result when the intermediate stimulus is more similar in size to one of the extremes than to the other. All of the experiments herein reported employed asymmetrical training sets: this appeared to be the first time in the history of the intermediate size problem that such sets were studied. If asymmetry (which can be of varying degrees) acts to change the values of the constants, a y value of .30 is possible in the highly asymmetrical cases studied in Exp. II and III and in some of the cases of Exp. I (in that study, the considerable range of asymmetry together with the small N per condition has made the differential effect of degrees of asymmetry difficult to evaluate) while it may be .60 for the Stevenson and Bitterman investigation where only symmetrical sets were employed. Further study is required to determine the role of symmetry in the establishment of the values of the x and y constants.

The gradual shifting of response to the formerly negative stimulus as revealed by Exp. III, the second new dimension revealed by this study, is significant for the same reason as the first effect: it is not readily conceivable as to how this phenomenon could be either expected or deduced by theoretical conceptions other than the ratio theory. These two new dimensions of the problem supplement the body of data on the intermediate size problem that have clearly revealed the inadequacy of absolute and relational theories. At the current time, the only approach capable of dealing with all of the data would seem to be the ratio theory.

Ability to abstract the concept of intermediate size apparently had no influence on behavior at the age level investigated. Either the fact of conceptualization has no necessary implications for choice performance or else other factors operated to override the influence of such abstraction. At any rate, the ability to abstract the concept would appear to be an irrelevant talent in connection with the discrimination of intermediate size by 4- and 5-yr.-old children where response can be explained on the basis of the ratio theory without reference to higher mental processes.

REFERENCES

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 565–621.

SPENCE, K. W. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, **31**, 247–271.

STEVENSON, H. W., & BITTERMAN, M. E. The distance-effect in the transposition of intermediate size by children. *Amer. J. Psychol.*, 1955, **68**, 274–279.

ZEILER, M. D. The ratio theory of intermediate size discrimination. *Psychol. Rev.*, 1963, **70**, 516–533.

(Received December 31, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 596–600

PROTECTIVE SELF-PACING DURING LEARNING[1]

WILLIAM H. SAUFLEY, JR.[2] AND INA McD. BILODEAU

Tulane University

4 groups of Ss practiced reversed printing under standard (no mention of duration) or experimental (duration of practice specified) instructions. Experimental instructions had no effect in the 1st practice period of naive Ss who had no recent past experience of massed practice on the task. Experimental instructions did affect performance of other groups in a 2nd period following 12 min. of continuous practice and a 10-min. rest. The Ss who expected 20 sec. of practice printed faster than control Ss given standard instructions, while Ss expecting 20 min. of practice were slower. Time-set thus can modify behavior in learning as well as in work, and recent massed practice is of primary importance in controlling self-pacing.

Among the accounts offered of the effects of distribution of practice and work-loading is that *S* adjusts his output to reduce the amount of reactive inhibition accrued (Bilodeau, 1952; Reynolds & Bilodeau, 1952)—i.e., that when conditions favor *S*'s accruing large amounts of response inhibition, *S* inhibits the inhibition-producing responses. Supporting evidence for the existence of such pacing adjustments can be found in a number of studies in which time- or amount-sets are induced by instructions about how long *S* must work, or how much work *S* must do (Bills & Brown, 1929). The strength of response at the start of practice depends on how many Rs *S* expects to make (Ross & Bricker, 1951) or how long he expects to practice (Bilodeau, 1953).

There are, however, two recent exceptions, both testing for conservation effects during learning; and it is ordinarily in learning experiments that protective self-pacing has been invoked. Noble (1953) failed to find evidence of amount-set with the Complex Coordination Test—a failure he was inclined to interpret as reasonable in view of the low work-loading of the task. Later, however, Noble and Cieutat (1958) found no conservation on the Two-Hand Coordination Test, again with amount-set, and questioned the generality of instructionally-induced conservation behavior.

Rather than accept the variable as effective after, but not during, learning, the present writers assign previous failures to three factors that affect how *S* interprets instructions aimed at provoking conservation sets. Each of these contributes to the meaning the instructions have for *S*. (*a*) Lack of prior experience of R-produced inhibition on the experimental task: prior experience gives *S* knowledge of what he is conserving himself against as opposed to the *S* who has not suffered decrement. (*b*) The ambiguity of amount sets: an *S* who knows that so many units must be completed before he leaves may be moved by two conflicting tendencies—to slow his rate of response to protect himself against the painful effects of responding rapidly and to speed up to escape to his own affairs the sooner. (Responses that vary in magnitude rather than speed are not subject to this criticism.) (*c*) Using *E*-paced rather than *S*-paced tasks: on an *E*-paced task, again, instructions may provide *S* with cues appropriate to a low rate of responding, but *E* also offers strong and contradictory cues to the rate of responding *S* should maintain.

---

[1] This paper is based on a thesis, "Time Set in a Learning Task," submitted by the first author to the Graduate School of Tulane University in partial fulfillment of the requirements for the Master of Science degree.

[2] Now at Northwestern University.

Noble and Cieutat may have found their instructions ineffective because: (*a*) their *S*s were naive to the task when instructed; (*b*) *E* used amount-set instructions; or (*c*) the Two-Hand Coordination Test is an *E*-paced task. An alternative explanation is that tasks with a heavy learning component are not subject to conservation pacing. To check this last possibility, the present writers aimed at conditions they consider favorable to demonstrating conservation by combining a self-paced task, time-set instructions, and *S*s with some experience of massed practice on the task. The task chosen was reversed printing, for which both sizable learning and decremental effects are known (Kientzle, 1946; Kimble, 1949). The major groups of the present experiment received a lengthy period of continuous practice under standard instructions, a rest to allow dissipation of temporary decrement, and a final period in which instructions about the duration of the forthcoming practice varied from group to group. As a test of the effects of experience on conservation, one group of naive *S*s was told the duration of the forthcoming first practice period.

METHOD

Subjects.—The *S*s were 137 students taking the introductory course in psychology at Tulane University. They were run in intact groups, all members of a class receiving the same treatment. There were four treatments, with 23–45 *S*s per treatment. No control of sex or preferred hand was attempted. An additional 9 *S*s were discarded for failing to mark the time intervals properly.

Apparatus.—The *S* received a booklet of seven sheets of standard four-to-the-inch graph paper. A page of the booklet contained about 700 small squares in a 24 × 29 matrix, outlined by heavy lines. There was no cover sheet. The *S*'s task was to print the letters of the alphabet, one letter per small square, from left to right and from A to Z, but upside down and backwards. The *S* started in the upper left hand corner of the top sheet.

Instructions.—Standard instructions: The task was explained and illustrated with a half dozen examples from the beginning of the alphabet, and the letters H, I, N, O, S, X, and Z pointed out as appearing unchanged in reversed printing. The *S*s were told to print as fast as possible, one letter per square within the matrix, proceeding directly to the next line when one line was completed. The *S* was to use his preferred hand, start at the "go" signal and continue until *E* said "stop," skipping a square between letters when *E* said "skip" and turning a page when *E* said "turn."

Experimental instructions: The *E* added to the standard instructions a statement that *S* was to perform the task for a given length of time.

Procedure.—To test for conservation in inexperienced *S*s, one group of *S*s, Group I-20 Min., received a single practice period of 20 min. duration and was told at the outset how long the period was to be. A control comparison was provided by three other groups of *S*s given a first practice period under standard instructions.

To test for conservation in experienced *S*s, the latter three groups received standard instructions before a first period of 12 min. practice, a constant 10-min. rest, and a final period before which the treatment differences were introduced. The *S*s of the control treatment, Group F-C, were given standard instructions again, with no mention of the length of practice. Group F-20 Sec. was told in addition that the practice period would be 20 sec. in duration and Group F-20 Min. that they would practice for 20 min. The two experimental groups of final practice thus contrast the effects of telling *S* he will have a very brief trial with the effects of telling him he will have a very long session.

TABLE 1

SUMMARY OF EXPERIMENTAL DESIGN

Group	N	Initial Practice Duration		Final Practice Duration	
		Stated	Actual	Stated	Actual
I-20 Min.	23	20 min.	20 min.	—[b]	—[b]
F-20 Sec.	45	—[a]	12 min.	20 sec.	30 sec.
F-20 Min.	36	—[a]	12 min.	20 min.	16 min.
F-C	33	—[a]	12 min.	—[a]	12 min.

[a] No reference to duration of practice.
[b] No final practice period.

Table 1 summarizes the design and gives the actual lengths of the final practice period and the number of *S*s in the several groups. The actual durations of final practice differed from those given in *E*'s instructions so that two 15-sec. scores would be available for Group F-20 Sec. and so that the entire experiment (including collecting the booklets) could be fitted conveniently into the 50-min. class period. During the rest, *S*s turned their booklets face down, while *E* showed them slides testing perception under conditions of impoverished cues.

Response measures.—Both practice periods were broken up into 4-min. scoring blocks by having *S* start a new page to mark the interval. In addition, *S* skipped a space every 15 sec. for the first minute and every 60 sec. for the other 3 min. of the first 4-min. scoring block.

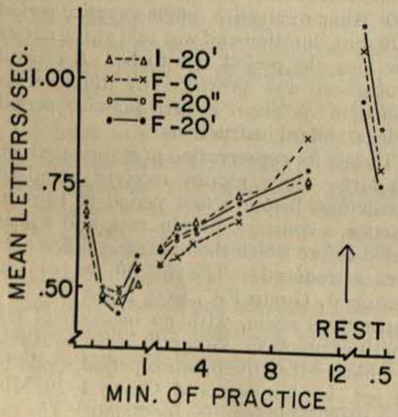


FIG. 1. Mean letters printed per second for scoring intervals of initial and final practice.

RESULTS AND DISCUSSION[3]

Figure 1 presents the mean number of letters printed per second in 12 min. of initial practice and 30 sec. of final practice.[4] To emphasize critical portions of the performance curves, the time intervals for final practice and for the first minute of initial practice have been drawn to four times the scale of the other intervals; the discontinuity between Min. 1 and 2 indicates this change in scale. Responses are plotted on the same scale throughout with data points located at the midpoints of the scoring interval over which they were averaged.

The data of initial practice support the original hypothesis about inducing pacing in naive *S*s: Group I-20 Min. (open triangles) is comparable to the other three groups. Note particularly that responses in the first minute are not depressed below the level of the three final practice groups. The latter three groups seem well matched in initial practice, particularly Groups F-20 Sec. and F-20 Min. The only sizable discrepancy between groups is in the last 4 min. of initial practice where Group F-C, previously slightly depressed, has a large jump in rate of responding. Variances increased very slightly within the first minute of initial practice, but were stable within and between groups thereafter.

---

[3] Tables of the data summarized in this report have been previously presented by Saufley (1961).

[4] Initial practice beyond 12 min. for Group I-20 Min. and final practice beyond 30 sec. for Groups F-C and F-20 Min. are not presented, as irrelevant to the main hypotheses.

Final practice shows the effect of conservation-inducing instructions on sophisticated *S*s: Group F-20 Min. is lowered, while Group F-20 Sec. is raised above the level of the control group. The difference between the two experimental groups is significant beyond the .001 level in the first 15 sec. of final practice, $F\ (1, 79) = 19.2$.[5] That the control group, F-C, falls between the experimental groups is in line with an earlier suggestion of a contrast effect for *S*s who expect a very brief trial after a long session (Bilodeau, 1953). The contrast effect must remain a suggestion, however, since the difference between Groups F-20 Sec. and F-C in the first 15 sec. does not reach even the .10 level, $F\ (1, 76) = 1.67$.

The present experiment was performed to verify that learning tasks are susceptible to self-conservation by combining three conditions considered favorable to protective self-pacing: some recent history of decrement; an *S*- rather than an *E*-paced task; and time- rather than amount-set instructions. Since the last two were not varied independently, their relative importance has not been assessed in the present design. The experiment does demonstrate that performance during learning can be influenced by the duration of practice expected, and that past practice is important in producing the effect.

A finding consistent for all groups and both practice sessions is the striking performance decrement in the early segment of the curve. Most of the decrement takes place in the first 30 sec. of practice, but it is not until Min. 5–8 that prerest performance slightly exceeds the starting rate. The initial negative slope, though exaggerated by our 15-sec. trials, is characteristic of continuous reversed printing. Less dramatic decrement in 30-sec. trials has been tentatively assigned to *S*'s being confused (Kimble, 1949; Wasserman, 1951) and to a short-term inhibitory factor (Zeaman & Kaufman, 1955). The present *E*s, suspecting that the decrement was an artifact of a serial-position effect (with all *S*s in phase at the start of practice and beginning with A), varied the starting letter in a later study (Saufley & Bilodeau, in press); the data are not in line with a simple serial position effect, but do suggest that the initial dip depends upon *S*'s knowing with which letter he will begin, rather than the particular letter with which he does begin.

REFERENCES

Bills, A. G., & Brown, C. The quantitative set. *J. exp. Psychol.*, 1929, **12**, 301–322.

Bilodeau, E. A. Decrements and recovery from decrements in a simple work task with variation in force requirements at different stages of practice. *J. exp. Psychol.*, 1952, **44**, 96–100.

Bilodeau, I. McD. Performance of an effortful task with variation in duration of prior practice and anticipated duration of present practice. *J. exp. Psychol.*, 1953, **46**, 146–153.

Kientzle, M. J. Properties of learning curves under varied distributions of practice. *J. exp. Psychol.*, 1946, **36**, 187–211.

Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.

Noble, C. E. Amount set and the length-difficulty function for a self-paced perceptual-motor skill. *J. exp. Psychol.*, 1953, **46**, 435–439.

---

[5] This comparison was settled before the experiment was performed, on the rational grounds that instructions will have their simplest effect early in practice (Bilodeau, 1953) and that Group F-20 Sec. would likely outperform Group F-C.

NOBLE, C. E., & CIEUTAT, V. J. Time set in two-hand coordination. *Percept. mot. Skills*, 1958, **8**, 82.

REYNOLDS, B., & BILODEAU, I. McD. Acquisition and retention of three psychomotor tests as a function of distribution of practice during acquisition. *J. exp. Psychol.*, 1952, **44**, 19–26.

ROSS, S., & BRICKER, P. D. The effect of an amount set on a repetitive motor task. *J. exp. Psychol.*, 1951, **42**, 39–43.

SAUFLEY, W. H., JR. Time set in a learning task. Unpublished master's thesis, Tulane University, 1961.

SAUFLEY, W. H., JR., & BILODEAU, I. McD. Anticipatory rehearsal in reversed printing. *Percept. mot. Skills*, in press.

WASSERMAN, H. N. The effect of motivation and amount of prerest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

ZEAMAN, D., & KAUFMAN, H. Individual differences and theory in a motor learning task. *Psychol. Monogr.*, 1955, **69**(6, Whole No. 391).

(Received January 11, 1963)

ERRATA

In the article by Cotton, Jensen, and Lewis (*J. exp. Psychol.*, 1962, **63**, 555–562) Equation 5 on page 559 should read:

$$H = 1 - B(10^{-iN})$$

Also, the last line of the left-hand column on page 560 should read:

distribution at ±2.5 σ_{diffs} from its mean,[2] the

In the supplementary report by Hammerton (*J. exp. Psychol.*, 1963, **66**, 108–110) references were made (p. 108, 110) to a paper by Fleishman and Parker (see *J. exp. Psychol.*, 1962, **64**, 215–226) which were inappropriate and which led to some misleading comparisons between the two studies. First, Fleishman and Parker studied retention on a task *more difficult* than Hammerton's. Secondly, they did, in fact, use acceleration controls. Finally Fleishman and Parker did not ascribe overlearning to their *S*s, although the initial learning period for their *S*s was much longer than for Hammerton's *S*s. Thus, as far as Hammerton's study goes, the results actually confirmed the Fleishman and Parker conclusion that, discounting an initial warm-up trial, retention of a continuous-control perceptual motor skill is remarkably high. The additional value of Hammerton's study is in showing that overlearning has the effect of reducing even this first trial performance decrement.

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 601–608

ROLE OF AWARENESS IN LEARNING AND USE OF CORRELATED EXTRANEOUS CUES ON PERCEPTUAL TASKS[1]

CHARLES W. ERIKSEN AND LEE DOROZ

University of Illinois

In 2 experiments Ss were presented with ostensibly perceptual tasks in which extraneous but correlated perceptual cues were deliberately provided in order to determine whether Ss would utilize these cues in guiding their responses, and further, whether they would do so without awareness as assessed by S's inability to verbalize the nature of the cue and its relation to his performance. Across experiments approximately 30% of the Ss were aware of the cue and the majority of them used it as a guide in their judgments. There was no evidence in either experiment that unaware Ss, including those who could name the cue only in response to a forced-choice question, had learned or employed the cue in their judgments.

While there has been extensive research on learning without awareness during recent years the learning materials and tasks employed have been quite restricted. The major effort has been on verbal conditioning, e.g., Greenspoon (1955) and Taffel (1955), although some scattered attempts to demonstrate operant conditioning of nonverbal responses without awareness have appeared (Hefferline, Keenan, & Harford, 1959; Paul, Eriksen, & Humphreys, 1962; Verplanck, 1956). Perhaps the increasing negative evidence (Dulany, 1961, 1962; Levin, 1961; Spielberger, 1962) for learning without awareness in adult *S*s is in part a function of the restricted types of learning materials or situations that have been investigated.

If adults do learn without awareness it would seem most likely to occur in relatively simple situations, situations requiring a minimum of abstraction or concept formation. A task where *S* has the opportunity to learn to use an extraneous or nonsalient perceptual cue as a guide for his behavior would seem to meet these requirements.

There are several reasons why one might expect such tasks to reveal learning without awareness. First, perceptual theories, most notably that of Brunswik (1943) have stressed the sensitivity of the perceptual system to contingencies among cues. A second reason is found in the criticism of experiments where an uncontrolled extraneous but correlated cue may have guided *S*'s responses. This particular criticism frequently has been leveled at ESP experiments where critics have suggested that an uncontrolled cue was the basis of *S*'s correct responses even though *S* may not have been aware he was utilizing this cue.

In the two experiments reported below *S*s were confronted with ostensibly perceptual tasks in which extraneous but correlated perceptual cues were provided in order to determine whether *S*s would utilize these cues in guiding their responses, and further, whether they would do so without awareness as assessed by *S*'s inability to verbalize the nature of the cue and its relation to his performance.

---

[1] This research was supported by Mental Health Grant M-1206, National Institutes of Health, Public Health Service.

EXPERIMENT I

Method

Procedure.—The *S*s were run individually. Upon appearance at the laboratory *S* was informed he was to participate in an experiment on unconscious perception. He was further told that he would be shown, one at a time, a list of pleasant and unpleasant words but the exposure of the words would be too rapid for him ever to feel sure he had really seen them. It was explained, however, that *E*s had good reason to believe that while he might not consciously recognize the word he nevertheless possessed the capacity to make correct emotional discriminations about the nature of the word shown. Thus while he would not be able to identify the word he should get a feeling whether it was pleasant or unpleasant and that this was the judgment that he was to make after each word exposure. In making his judgment as to whether a pleasant or unpleasant word had been exposed *S* was instructed to use whatever cues he felt might guide him in his decision. He further was told that he would be corrected after each decision as to whether he was right or wrong and that while he was not expected to be too good at the task at the beginning he should after a number of corrected trials become more proficient at making the emotional judgment.

The true nature of the experimental situation was quite different from that presented to *S*. Instead of presenting pleasant and unpleasant words, 20 two-syllable nonsense syllables constituted the stimuli. The list was replicated once with the replication being identical to the original list of 20 with the exception that in the replicated syllables either the first or the last letter had been deleted and replaced by a dash. For half of the *S*s the occurrence of an altered nonsense syllable constituted a "pleasant" word whereas for the remaining *S*s it was called "unpleasant."

The stimuli were presented to *S* in a tachistoscope. Each of the 40 stimuli was presented a maximum of five times for a total of 200 corrected presentations with the order of occurrence of the stimuli determined by a table of random numbers with the restriction that each of the 40 stimuli occur in successive blocks of 40 presentations and that each successive block of 20 presentations contains 10 altered and 10 unaltered stimuli. A further restriction eliminated runs of greater than three.

The first 10 min. of the experimental session was spent determining an exposure duration for the stimuli to be used with *S*. The *S* was presented the capital letters A or T in random sequence in the T-scope and required to make a forced-choice discrimination as to which letter had been shown. By systematically varying the exposure durations a level was obtained for each *S* where he correctly identified approximately 75% of the presentations of A and T. This duration level was then used for the presentation of the nonsense syllables during the experimental judgments. Preliminary work had shown that this threshold determination resulted in an exposure level for the nonsense syllables that made the detection of the dash possible but not too obvious.

Following the judgments *S* was asked in order the following questions:

> 1. You became fairly good at this task. Could you tell me what kind of a cue or what the basis of your judgment was? Did you just have a feeling about it or did the words look different or what? (If *S* did not spontaneously verbalize the correct cue at this point he was asked the following questions.)
> 2. Actually there was a cue present; a basis by which you could tell whether there was a pleasant or an unpleasant word presented—something about the way it looked. Did you notice what this cue was? (If *S* failed to verbalize the correct cue at this point, he was told the nature of the cue and asked whether it had been associated with the pleasant or the unpleasant category.)

These questions enabled us to get around the ambiguity that typically characterizes interrogation for awareness and at the same time, by working down to a forced-choice guess, permitted a more explicit statement of just what *S*'s level of awareness was.

Apparatus.—A Dodge-type mirror tachistoscope was employed having an illumination in the adapting field of .5 apparent ft-c and a brightness in the stimulus field of .3 apparent ft-c. A cross in the adapting field was employed as a fixation point and the stimuli were presented 6′ above the fixation point and centered upon it in the horizontal dimension. The stimuli were typed in capital letters on an electric typewriter and the dash employed at the beginning or end of the altered nonsense syllables was the dash from the keyboard of the typewriter which replaced either the initial or the final letter of the nonsense syllable.

Subjects.—Thirty-eight male and 39 female undergraduate students enrolled in the introductory psychology course at the University of Illinois were *S*s. As part of the course requirements *S*s are required to serve a maximum of 5 hr. as experimental *S*s during the semester. Half of the males and 20 of the females were assigned to the condition where the altered syllable was called pleasant by an essentially random procedure based upon the order of the appearance of *S*s for the experiment.

Results

Twenty-seven correct responses out of a block of 40 trials were adopted as a criterion of learning for individual *S*s, χ^2 (1) = 4.9, $p < .05$. Without knowledge of the learning performance, two raters working independently evaluated *S*s' responses to the interrogation that took place at the end of the experiment and classified *S*s into aware and unaware groups. The raters further specified which of the first two questions elicited evidence of awareness and for unaware *S*s they further subdivided this group into those who responded correctly to the forced-choice question as opposed to those who did not. Agreement between the raters was perfect.

Table 1 presents the number of *S*s in each awareness classification by sex and by whether the cue was designated pleasant or unpleasant and also shows the number who met the criterion for learning.[2] Since it is apparent from Table 1 that there are no appreciable differences due to sex or to cue designation, these distinctions have been abandoned in subsequent analyses.

A fourfold chi square comparison of the number of *S*s satisfying the learning criterion in the aware and unaware groups yielded a value of 41.57 significant beyond the .0001 level. Of the 23 aware *S*s, 2 met the criterion within the first block of 40

TABLE 1

NUMBER OF *S*S BY SEX AND CUE DESIGNATION IN THE THREE AWARENESS CLASSIFICATIONS WHO SATISFIED THE CRITERION OF LEARNING

	Aware		Unaware			
	Learners	Nonlearners	Correct on F-C Question		Incorrect on F-C Question	
			Learners	Nonlearners	Learners	Nonlearners
Pleasant cue						
Male	3	1	0	9	1	4
Female	4	2	0	5	0	8
Unpleasant cue						
Male	6	1	0	5	0	7
Female	5	1	1	7	1	4
Total	18	5	1	26	2	23

---

[2] The data for two *S*s do not appear in Table 1. Both raters judged these two *S*s as having a correlated hypothesis, namely, that the cue stimuli appeared shorter. Since replacing a letter at the beginning or the end of the nonsense syllable with a dash did lead to the appearance of a shorter word during brief tachistoscopic exposures, this hypothesis would be a valid guide to correct discrimination. Both of these *S*s met the criterion for learning but due to the small number of *S*s in this group further analyses involving them were not feasible.

trials and 10 had satisfied the criterion within 160 trials. Sixteen of the aware *Ss* either spontaneously verbalized their awareness during the course of the experiment or in response to the first question during the interrogation.

The number of unaware *Ss* who correctly attributed the presence of a cue to the pleasant or unpleasant designated stimuli on the forced-choice question closely approximates a chance division. However, to determine whether correct response on the forced-choice question represented a minimal degree of awareness, the two subdivisions of the unaware *Ss* were compared on the number of correct responses made during the last 40 trials. The means were 19.26 and 20.92 for the correct and incorrect choice subgroups, respectively. The direction of the difference is actually in favor of more correct responses for *Ss* making the incorrect choice to the forced-choice question. However a *t* test of the difference yielded a value of 1.55 which with $df = 50$ is not significant.

Since the exposure duration in the tachistoscope of the stimuli was determined individually during the first part of the experimental session, it is possible that aware *Ss* may have been those who had longer exposure durations and were therefore more likely to detect the presence of the dashes before or after the cue stimuli. To check this possibility the mean duration level of exposure for *Ss* in the aware and unaware groups was computed. The respective mean values were 59.3 and 59.1 msec. A test of this difference gave a *t* value of less than 1.

The results of this experiment clearly demonstrate that some *Ss* will detect and use a correlated cue as a basis for their responses in a perceptual task. The results are equally clear in showing that *Ss* who are unable to verbalize the nature of the cue and its relation to the stimuli designated pleasant or unpleasant did not learn.

It is of course possible that our failure to find any evidence of *Ss* learning to use a correlated perceptual cue without awareness may have been a function of the specific task employed, the set of the *Ss*, or other undetected factors in the experimental situation. Accordingly a second experiment was carried out employing a different type of perceptual task. This experiment was suggested from a study by Brunswik and Herma (1951) in which they reported that they had induced weight illusions in *Ss* by a correlation of position cues with differential weight magnitudes. They further reported that these cues produced significant illusory effects in judgments even though *Ss* were unaware of the relationship between position and weight magnitude.

EXPERIMENT II

Method

Procedure.—The *Ss* again were run individually. The *S* was informed that he was participating in an experiment upon the effect of color on the apparent length of lines. He was told that he would be shown a number of pairs of lines where the members of the pair were of different colors and of different lengths. He was to judge whether the line on the left or the line on the right appeared the longer. He was further instructed that on some of the judgments the difference would be fairly obvious whereas on the other trials the difference would be difficult to detect. However he was always to make his judgment or choice in terms of which line appeared the longer. The *S* was then seated at a table that was divided down the center by a vertical plywood screen that shielded *E* from *S*'s view. An aperture in the screen was used for stimulus presentation to *S*.

The stimuli for the learning session consisted of a set of 60 8½ × 11 in. plain white cards upon which were drawn two parallel vertical lines 1 in. apart and located approximately in the center of the card. There was a vertical difference of about ¾ in. between the corresponding terminal points of the separate lines with the exact difference dependent upon the line lengths employed. The lines were drawn with a lettering set using colored inks. The four colors red, green, blue, and yellow were used and the two lines on any one card were always of a different color. On 52 of the 60 cards one of the lines was always 3 in. in length and the comparison line was either 3⅛, 3¼, or 3⅜ in. On the remaining six cards the two lines were equal in length, three of these pairs being 3 in. and three, 3⅜. The various line length combinations were counterbalanced with right-left position and also with color except for the experimental treatment to be described below. The stimuli were placed in a random order that remained constant for all *S*s.

Two experimental groups were run, a yellow and a blue group. For the yellow experimental *S*s a yellow line was always the longer line of any pair in which it appeared. Counterbalancing of other colors in all possible combinations of lengths thus required that red, green, and blue each be the longer line 12 times, the shorter line 18 times, and equal in length 4 times. Yellow, however, was the longer line 18 times and never appeared as shorter or equal to a comparison line. This set of 60 cards was presented twice to each *S* thus providing *S* with 36 opportunities to learn that yellow always was associated with longer lines.

Following the learning series of judgments, *S*s were given a 4-min. rest period before judging the test series. For the test series 18 cards were added upon which yellow was presented equal in length to one of the other colors. Nine of these cards had equal line pairs of 3 in. and the other nine of 3⅜ in. Red, green, and blue were each used six times and position counterbalanced. This test series of 78 cards was also presented twice with the 18 additional cards placed randomly throughout the 60, placement held constant for all *S*s. Thus in the test series there were 36 cards where yellow was actually equal in length to the other colored line and where *S* was forced to make a choice as to which line was longer.

The above procedure was identical for *S*s in the blue experimental group except that blue was substituted for yellow as the correlated color cue.

Because of the possibility of color preferences in *S*s and of real effects of color upon apparent line length it was necessary to employ two control groups (yellow and blue) as a means of assessing whether the experimental *S*s actually employed the color cue during the test judgments. The *S*s in both control groups were treated identically with their corresponding experimental *S*s except the correlated color cues were removed for the control *S*s by having each color during the learning series appear as longer and shorter than the comparison color equally often.

Following the test series judgments *S*s in all four groups were interrogated with the following questions.

1. How did you go about making close decisions?
2. Were there any other cues that you found helpful, or did anything else seem to influence your judgment in any way?
3. How about color? Was there any particular color that seemed always to be associated with a longer line?
4. If you were forced to choose one particular color that was longer every time that it appeared what would that color be?
5. Did you think about this at all while watching the cards or is it just on looking back that you think (*color*) was always longer or is this just a pure guess?
6. (To test *S*'s reaction *E* then said) Well you're wrong. Actually (a different color than that stated by *S*) was always longer. (This statement, however, elicited no argument from any *S* since even those *S*s subsequently classified as aware apparently were not highly confident of the contingency they had detected.)

Subjects.—Forty-two males and 34 females were again drawn from the introductory psychology course. The same sampling procedure from the pool of *S*s was employed in obtaining *S*s for each of the four groups. It was originally planned to have 30 *S*s in each of the two experimental groups and 14 in each control group proportionately distributed between male and female. Due to failure of a few *S*s to appear for their experimental appointment the yellow cue experimental group contained 25 *S*s and the blue, 27. The yellow control group contained 13 and the blue group, 11.

Results

The *S*'s interrogation was evaluated independently by two raters without knowledge of *S*'s performance on the color judgments. The *S*s were classified as aware if they correctly verbalized the contingent relation between color and line length in response to any one of the first three questions. Unaware *S*s were subdivided further based upon whether or not they had chosen the correct color in response to Forced-Choice Question No. 4. Agreement between raters was perfect.

To insure that the differences in length employed during the learning series were indeed great enough for *S* to learn the color cue, judgmental accuracy for the learning series was computed for each *S*. The average accuracy was 96% with only one *S* falling below 90%. She was discarded from the following data analyses. An analysis for sex differences was also carried out and since no significant nor suggestive differences were noted the sex classification was disregarded in subsequent analyses.

Table 2 shows the number of *S*s in the two color cue groups that were classified aware as well as a breakdown of the unaware *S*s into those who named the correct color in response to the forced-choice question. The mean number of illusory responses during the test judgments is also given in Table 2 for the various subgroups and the two control groups. As can be seen, for both color cues the aware *S*s gave more responses in the direction of the illusion than did the control or unaware *S*s. For the yellow cue the unaware *S*s, irrespective of whether or not they named the correct color in response to the forced-choice question, did not differ appreciably from the performance of the control *S*s. For the blue cue the control group actually had a higher mean than either of the two subgroups of unaware *S*s.

The significance of these differences between means was evaluated by a simple analysis of variance employing the three subgroups of the experimental *S*s and the corresponding control group but performed separately for the yellow and the blue cue. For the yellow cue $F\ (3, 36) = 6.59$ for between-groups which is significant beyond the .01 level. The t tests of the mean differences using as the variance estimate the error term from the analysis of variance ($s = 3.4$) revealed that aware *S*s differed from the control as well as from the two subdivisions of the unaware *S*s at a confidence level beyond .01. The *S*s who were unaware of the cue and who made correct or incorrect choices did not differ from the control group nor from each other (all t values less than 1).

A corresponding analysis of variance for the blue color cue failed to

TABLE 2

Mean Number of Illusory Direction Judgments by Awareness Classification

	Control	Aware	Unaware	
			Correct on Forced-Choice Question	Incorrect on Forced-Choice Question
Yellow cue				
N	13	9	7	11
M	13.8	19.8	13.7	14.4
Blue cue				
N	11	5	7	13
M	22.3	25.0	19.6	20.9

show significant differences between the four groups, $F(3, 32) = 2.12$. The direction of the differences was consistent with the results from the yellow cue group.

A comparison of the means for the yellow and blue control groups in Table 2 also revealed that when *S*s are presented with two equal length lines and asked to judge which is longer the color of the lines makes a difference in their judgment. A *t* test of the difference between the means for the two control groups yielded a value of 2.84 significant beyond the .01 level which indicated that *S*s tended to judge blue lines longer more often than yellow lines.

DISCUSSION

The results of the two above experiments are unequivocal in showing that a certain proportion of *S*s will learn and use a correlated extraneous cue as a guide for their behavior in an experiment. The results are also quite clear in showing that this learning and use of the extraneous cue occurs only among *S*s who are able to verbalize clearly the nature of the cue.

It has long been known in the field of psychophysics that some *S*s can make reliable discriminations while reporting that subjectively they are purely guessing (Pierce & Jastrow, 1884). It was partly on the basis of these phenomena that we included the forced-choice question in our interrogation in the above experiments. We felt that those *S*s making the correct choice or response to this question might represent a minimal level of awareness of the existence of the cue. Also the use of such a question avoids some of the subjective problems in specifying awareness. It was clear from the results in both experiments that if a correct forced-choice answer did indicate a minimal degree of awareness it was not sufficient to affect *S*'s behavior on the test judgments.

The negative evidence for learning without awareness in the present experiments is reinforced by our finding that a significant proportion of the *S*s learned the cues with awareness and that this learning did affect their test judgments or performance on the experimental tasks. The fact that approximately 30% of the *S*s became aware and learned the cue during the experiments indicates that the tasks were sensitive enough, first, to produce learning of the cue, and secondly, to provide enough leeway for *S*s to learn without awareness if such a process had been operating in the present experimental arrangements.

The fact that only approximately 30% of the *S*s learned and became aware of the correlated cue in the above studies suggests that human adults are rather insensitive to contingencies in their environment unless they are programed through instructions or prior induced set to look for these contingencies. Casual conversation with our *S*s following the experimental session suggested that aware *S*s were predominantly those who came to the experiment with a certain degree of suspicion or who felt there was some gimmick involved and were actively looking for it in the experimental procedure. This insensitivity to contingencies in the environment among adults is consistent with findings from other experiments in our laboratory. In a recent study (Paul et al., 1962) we found that, again, only approximately 30% of the *S*s detected the correlation between the advent of cool air reinforcement in a heat-humidity chamber and an operant response of their own. Perhaps children, having different expectancies and sets than adults, might be more sensitive to such contingencies.

REFERENCES

BRUNSWIK, E. Organismic achievement and environmental probability. *Psychol. Rev.*, 1943, **50**, 255–272.

BRUNSWIK, E., & HERMA, H. Probability learning of perceptual cues in the establishment of a weight illusion. *J. exp. Psychol.*, 1951, **41**, 281–290.

DULANY, D. E., JR. Hypotheses and habits in verbal "operant conditioning." *J. abnorm. soc. Psychol.*, 1961, **63**, 251–263.

DULANY, D. E., JR. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness: 1962*. Durham, N. C.: Duke Univer. Press, 1962. Pp. 102–129.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

HEFFERLINE, R. F., KEENAN, B., & HARFORD, R. A. Escape and avoidance conditioning in human subjects without their observation of the response. *Science*, 1959, **130**, 1338–1339.

LEVIN, S. M. The effects of awareness on verbal conditioning. *J. exp. Psychol.*, 1961, **61**, 67–75.

PAUL, G. L., ERIKSEN, C. W., & HUMPHREYS, L. G. Use of temperature stress with cool air reinforcement for human operant conditioning. *J. exp. Psychol.*, 1962, **64**, 329–335.

PIERCE, C. S., & JASTROW, J. On small differences of sensation. *Mem. Nat. Acad. Sci.*, 1884, **3**, 73–83.

SPIELBERGER, C. D. The role of awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness: 1962*. Durham, N. C.: Duke Univer. Press, 1962. Pp. 73–101.

TAFFEL, C. Anxiety in the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–502.

VERPLANCK, W. S. The operant conditioning of human motor behavior. *Psychol. Bull.*, 1956, **53**, 70–83.

(Received January 16, 1963)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 609–611

SUPPLEMENTARY REPORTS

COMPOUND NONSENSE-SYLLABLE STIMULI PRESENTED WITHOUT AN INTERVENING SPACE[1]

BARBARA S. MUSGRAVE
Smith College
ALBERT E. GOSS AND ELIZABETH SHRADER
University of Massachusetts

Musgrave's experiment on transfer of nonsense-syllable responses to a pair of nonsense-syllable elements presented separately to those elements presented simultaneously in Convergent, Associated-Nonassociated, and Divergent compounds was replicated in all essential features but one: the nonsense syllables were presented in compounds with no rather than with one intervening space between elements of the compound. In agreement with Musgrave's results, faster responding occurred under a Continued Singles control than under the Convergent compound. Responses under Associated-Nonassociated and Divergent compounds were successively slower than those under the Convergent compound.

Stimuli composed of two nonsense syllables each of which had previously been associated with the same response (convergent compound) elicited slower responses than either of the component syllables presented alone (Musgrave, 1962). In order to minimize the likelihood of generalization decrement due to a novel shift from one stimulus to two stimuli, *S*s given the convergent compound had had prior experience with two stimuli presented simultaneously. Thus, occurrence of two stimuli simultaneously during test trials was not novel.

Although grossly in agreement with one theorem based on Hull's (1943, p. 219) hypothesis of neural afferent interaction, the decrement in speed with convergent compounds did not seem consistent with implications of models which assume all-or-none associations between components or patterns and responses (Estes, 1959). However, another aspect of the changes from elements alone to elements in a compound conceivably unaffected by prior experience with two stimuli was the space between the two elements of the compound each of which had previously appeared singly. With newly-paired stimuli, the space may have continued to function as a new element to occasion generalization decrement.

In the experiment reported here, the two syllables of the convergent compound were presented simultaneously written from left to right with no intervening space. Thus, the space element was eliminated as a possible source of generalization decrement. Included in the present experiment were the other compounds and the control conditions of Musgrave's experiment. Thus, the present experiment constituted a partial replication of Musgrave's experiment.

Method.—The conditions, apparatus, materials, and procedure of the present experiment differed from those of Musgrave's in but two details: use of 8 rather than 16 *S*s in the Convergent condition, and, both during some training trials and during all test trials with compounds, presentation of the two stimuli simultaneously with no space between them. As before, *E* was a female and all *S*s were females.

Results.—Means and *SD*s of latencies for all eight conditions for the criterion (last training) trial and for each of the four test trials are presented in Table 1. As previously, latencies exhibited substantial heterogeneity of variance. For the analyses of variance, therefore, they were transformed into their reciprocals (speed of response evocations). Differences among conditions on the criterion trial were not significant, $F(7, 56) = 1.21$. In the analysis of variance on speeds during the test trials, the *F*s for Conditions, Trials, and Trials × Conditions were significant at less than .001. Because of the significant Trials × Conditions interaction, differences among conditions were assessed for each trial separately. These differences among conditions were significant for Trials 1 ($p < .001$),

---

[1] This research was supported in part by Contract Nonr 2691(00) between the Office of Naval Research and the University of Massachusetts.

TABLE 1

MEANS AND *SD*s OF LATENCIES ON THE CRITERION TRAINING TRIAL AND FOUR TEST TRIALS FOR ALL CONDITIONS

Cond.	Training		Test							
	Criterion Trial		Trial 1		Trial 2		Trial 3		Trial 4	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Continued Singles (Cont. S.)	1.025	.155	1.040	.148	.980	.179	.929	.118	.950	.200
Convergent (Conv.)	.982	.152	1.138	.118	1.056	.142	1.022	.188	.973	.257
Asso.-Nonasso. I (A-NA-I)	1.075	.314	1.262	.251	1.042	.195	.992	.237	1.078	.475
Asso.-Nonasso. II (A-NA-II)	1.028	.024	1.228	.168	1.166	.217	1.027	.142	1.086	.331
Divergent I (Div. I)	.983	.118	1.343	.185	1.584	.394	1.232	.185	1.114	.233
Divergent II (Div. II)	1.012	.142	1.177	.123	1.470	.380	1.202	.245	1.062	.241
New Singles (New S.)	1.016	.055	2.058	.984	1.292	.332	1.101	.167	1.138	.267
New Doubles (New D.)	1.049	.210	1.919	.377	1.276	.134	1.104	.131	.930	.100

2 ($p < .001$), and 3 ($p < .025$) but not for Trial 4. Table 2 summarizes the Duncan multiple-range tests of the significance of differences between all pairs of means of speeds on each of the four trials.

Additional analyses of variance for each condition separately indicated that, through the test trials, speeds increased significantly for all but the Continued Singles condition. Order of the syllables in the Convergent, Associated-Nonassociated, and New Doubles conditions had no differential effects.

Under the Convergent, Associated-Nonassociated, and Continued Singles conditions, all *S*s responded correctly on all test trials. For the Divergent conditions, either syllable which had been associated with the stimuli during training was scored as correct during the first two test trials. Under the Divergent I condition, on Trials 1 and 2, one *S* responded with both syllables and on Trial 2 another *S* gave only the single first letter of the correct syllable. Under both the Divergent II and the New Singles conditions, on Trial 1 one *S* gave only the single first letter of the correct syllable.

TABLE 2

SUMMARY OF DUNCAN RANGE TESTS

Cond.	Cont. S.	Conv.	Div. II	A-NA II	A-NA I	Div. I
			Trial 1			
Div. II	.05	—	—	—	—	—
A-NA II	.05	—	—	—	—	—
A-NA I	.05	—	—	—	—	—
Div. I	.001	.05	—	—	—	—
New S.	.001	.001	.001	.001	.001	.001
New D.	.001	.001	.001	.001	.001	.001
			Trial 2			
Div. II	.001	.05		—	.01	
Div. I	.001	.005		.05	.005	
New S.	.05	—		—	—	
New D.	.01	—		—	.05	
			Trial 3			
Div. II	.01	—		—	.05	
Div. I	.005	.01		.05	.01	
New S.	.05	—		—	—	
New D.	.05	—		—	—	

On Trial 2, the stimulus was followed by the correct response syllable. On Trials 3 and 4 only this syllable was scored as correct. Except for one *S* in the Divergent II condition, on Trial 3, all *S*s responded correctly on both trials.

Faster responses occurred under the Continued Singles condition than under the Convergent condition. In turn, responses under the Divergent condition were slower than those under the Convergent condition. Speeds under the Associated-Nonassociated conditions fell between those for the Convergent and Divergent conditions. The results, therefore, were markedly similar to Musgrave's. The space between syllables apparently did not introduce a new element or elements. Remaining as a possible explanation of the results of both the present and the Musgrave experiment is some yet undetermined difference between orientation to stimuli alone and to stimuli in compounds which introduces a new element or elements.

The slower speeds with the Convergent compounds may, in part, reflect longer time to read compound stimuli than single stimuli (Musgrave, 1962). But any increments in reading time of recognition responses to elements in compounds, as Musgrave and Cohen (1962) have suggested, may well be a consequence of kind of compound, number of elements in the compounds, and prior experiences with those elements. Whatever the bases for the observed difference in speeds, as with Musgrave's findings, it cannot be explained by present models which assume all-or-none, nonmediated associations between patterns or components and responses (Estes, 1959).

REFERENCES

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Musgrave, B. S. The effect of nonsense-syllable compound stimuli on latency in a verbal paired-associate task. *J. exp. Psychol.*, 1962, **63**, 499–504.

Musgrave, B. S., & Cohen, J. C. The effects of two-word stimuli on recall and learning in a paired-associate task. Paper presented at Eastern Psychological Association, Atlantic City, 1962.

(Received December 10, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 611–612

GOAL-BOX AND ALLEY SIMILARITY IN LATENT EXTINCTION

JERRY W. KOPPMAN AND G. ROBERT GRICE
University of Illinois

Equal latent extinction effects upon running time in a T maze were obtained with goal boxes which were distinctive from, and similar to the rest of the maze. It was suggested that the obtaining of greater effects with a distinctive goal box by Hughes, Davis, and Grice (1960) was attributable to the fact that it was larger than the nondistinctive goal box. The failure to obtain greater latent extinction effects, when the goal was similar to the alleys, was interpreted as evidence against the theory that the extinction of incentive motivation generalizes from goal box to alley on the basis of stimulus similarity.

Hughes, Davis, and Grice (1960) obtained a greater latent extinction effect upon running time in a T maze with a goal box which was distinctive from the rest of the maze, than with one identical to the maze alleys. The experiment was done to test a theoretical prediction based on the view that latent extinction is attributable to the extinction of incentive motivation in the goal box. Subsequent generalization of the extinction effect to earlier portions of the maze was assumed to be upon the basis of stimulus similarity. The outcome was opposite from the prediction. The authors, however, did not conclude that the experiment justified outright rejection of the theory under test. Their chief reservation was based on the fact that the distinctive goal box was larger than the nondistinctive one. In view of evidence suggesting that interfering effects may be greater in a large goal box, they suggested that goal-box size may have been responsible for the unpredicted outcome rather than distinctiveness per se. The present experiment is essentially a replication of the original experiment but with the distinctive and nondistinctive goal boxes of the same size and shape.

Method.—The *S*s were 48 experimentally naive male albino rats of the Holtzman strain. Half were 69 days old at the beginning of habituation; half were 77 days old. The apparatus was the same T maze described by Hughes, Davis, and Grice (1960) with two differences. In this experiment the two goal boxes were of the same size and shape, 14 × 4½ × 3½ in. The walls of the distinctive box were painted with ¾-in. alternating black and white stripes which continued across the floor. The floor was a ⅛-in. Masonite insert, rough side up. The nondistinctive goal box was of the same color and material as the rest of the maze. The distinctive goal box was enclosed in a compartment formed by two thicknesses of white cheese cloth. This isolation compartment was 23 × 18 × 14 in., and served to increase distinctiveness with respect to extramaze cues. A comparable 14-in. high curtain was also placed on the opposite side of the choice point to make the view from the choice point symmetrical.

The procedure was as described by Hughes, Davis, and Grice (1960). There were four groups: two latent extinction or experimental groups, one with the distinctive and one with the nondistinctive goal box; and a corresponding control group for each of these. Briefly, there were 24 acquisition trials, half free and half forced. There were 4 latent extinction test days. Each of these consisted of four 30-sec. placements in the empty goal box for the experimental groups, followed by a test trial and four retraining trials. On the fourth test day, animals were run to an extinction criterion rather than retrained. Treatment of the control groups merely omitted the placements in the empty goal box.

Results.—As in the previous experiment,

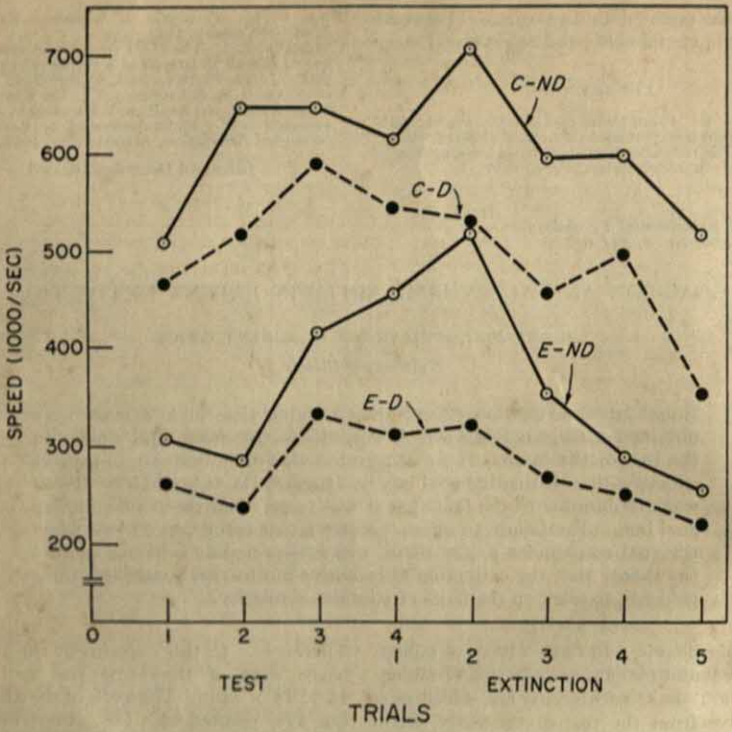


Fig. 1. Mean running speed on test trials and five extinction trials.

there were no significant differences between the groups in acquisition. Furthermore, as in the previous study, there was no evidence for the latent extinction effect upon choice behavior. There were only three reversals among all choice trials, no two by the same animal.

Running speed data on test trials and for five extinction trials are presented in Fig. 1. An inspection of these data indicates that the animals ran more slowly to the distinctive goal box, but that latent extinction effects were obtained about equally for the distinctive and nondistinctive conditions. These conclusions are supported by statistical tests. In an analysis of variance of test trial data, both major effects were significant, but their interaction was not—Latent Extinction, $F(1, 44) = 220.2$, $p < .001$; Distinctiveness, $F(1, 44) = 23.2$, $p < .01$; LE × D, $F(1, 44) = .001$. The Trials effect was also significant, $F(3, 132) = 4.3$, $p < .01$, but none of the interactions with Trials was significant. Analysis of the extinction data yielded essentially identical results.

The present findings fail to confirm that portion of the Hughes, Davis, and Grice (1960) result which included a significantly greater latent extinction effect with a distinctive goal box. This discrepancy tends to support the speculation of those authors that this finding was attributable to a greater interference effect with a larger goal box. This experiment does, however, confirm the original finding in failing to support the original prediction. Explanation of latent extinction in terms of generalization from goal box to maze suggests that the latent extinction effect should be greater with the goal box highly similar to the maze alleys. In view of these findings, and in view of analogous findings by Gonzalez and Diamond (1960) in a study of goal-box and alley similarity in an acquisition situation, there appears to be little present evidence for the view that incentive motivational effects occur early in a maze on the basis of stimulus generalization.

REFERENCES

Gonzalez, R. C., & Diamond, L. A test of Spence's theory of incentive-motivation. *Amer. J. Psychol.*, 1960, **73**, 396–403.

Hughes, D., Davis, J. D., & Grice, G. R. Goal box and alley similarity as a factor in latent extinction. *J. comp. physiol. Psychol.*, 1960, **53**, 612–614.

(Received December 7, 1962)

Journal of Experimental Psychology
1963, Vol. 66, No. 6, 613–614

EFFECT OF DURATION OF CONFINEMENT IN A NONBAITED GOAL BOX ON THE "APPARENT FRUSTRATION EFFECT"[1]

ANN W. ROBINSON AND KEITH N. CLAYTON

Vanderbilt University

32 rats received 51 trials training in a double runway maze. The experimental design was a 2 × 2 factorial simultaneously manipulating (a) duration of delay in 1st goal box (5 sec. vs. 5 min.), and (b) whether or not food was ever presented in the 1st goal box (experimental vs. control). During the subsequent 14-trial test for a frustration effect the experimental Ss received 50% reward in the 1st goal box. Only the 5-sec. delay experimental Ss ran faster in the 2nd runway on trials following nonreward than on trials following reward, supporting the hypothesis that the frustration effect is a transitory phenomenon.

In a study by Amsel and Roussel (1952) rats were trained to traverse a runway and enter a baited goal box. After a short delay access to a second runway and baited goal box was permitted. Following extensive training, when performance in the second runway was stable, nonreward was introduced in the first goal box with the result of faster running in the second runway. This increased vigor of responding, the *frustration effect* (FE), has been attributed to a motivational effect of nonreward. The Amsel-Roussel study, however, failed to detect a differential effect of duration of confinement in the unbaited goal box when this variable ranged from 5 to 30 sec. The present study extends the range of the duration variable to 5 min. to provide an improved test of Amsel and Roussel's hypothesis that FE would dissipate with time. Since the control conditions suggested by Wagner (1959) were utilized here, the study is essentially a replication of Wagner's, with the addition of experimental and control groups which were always detained for 5 min. in the first goal box.

Method.—The apparatus was patterned after the double runway used by Wagner (1959) with a single major modification, the second runway (a Hunter runway, Model 380) was shorter. The entire apparatus was 4½ in. deep and covered by glass. The lengths of the start box, first runway, first goal box, second runway, and second goal box were 11¼, 36¼, 12, 32¾, and 11¼ in., respectively. The start box and second goal box were 5 in. wide, elsewhere the maze was 3¾ in. wide. The start box, first runway, and first goal box had black ribbed rubber floors and black wood sides. The second runway and goal box had unpainted pressed wood floors covered with hardware cloth and glass sides with externally attached white cardboard. Three photocells were used to measure start and running times in the second runway. Opening the door of the first goal box started a 1/100-sec. electric clock which stopped when *S* interrupted an infrared beam 4¾ in. from the goal-box door. When the first clock stopped, a second clock (1/100 sec.) started which stopped when *S* activated the second photocell 21 in. from the first photocell. In similar fashion a third clock (1/10 sec.) started and was stopped by means of the photocell in the second goal box, 9 in. from the second photocell.

The *Ss* were 32 naive male albino rats of the Sprague-Dawley strain approximately 3 mo. of age at the beginning of preliminary training. The experimental design was a 2 × 2 factorial with eight randomly assigned *Ss* per cell. The two variables were: (*a*) duration of confinement in the first goal box (5 sec. vs. 5 min.), and (*b*) whether or not food was ever presented in the first goal box (experimental vs. control). The two subgroups of the short duration of confinement condition, therefore, were equivalent to Wagner's Groups C and N.

Preliminary and experimental training procedures were similar to those used by Wagner with the following exceptions: (*a*) 6 days of preliminary training were used, (*b*) during experimental training two trials were given daily, with the exception of the first day on which one trial was given, (*c*) the minimum intertrial interval was 2½ hr., (*d*) the reward in the first goal box, when given, was two 45-mg. Noyes food pellets and four pellets were given in the second goal box, (*e*) a total of 65 trials was given. For the experimental groups reward was always present in the first goal box for the first 51

---

[1] This study was supported in part by National Institute of Mental Health Grant M-5576 directed by the second author.

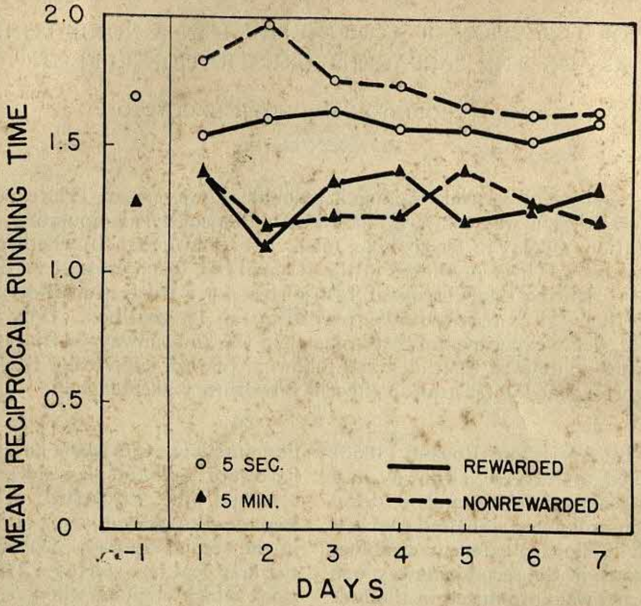


FIG. 1. Mean reciprocal Clock No. 2 time during the 7 days (Trials 52–65) for FE, experimental groups only.

trials. For the remaining 14 trials, the test period for FE, the experimental *S*s received 50% reward in the first goal box, each *S* being rewarded on only 1 of the 2 daily trials. Assignment of the rewarded trials was random with the restriction that half the *S*s in each experimental subgroup was rewarded on the first trial of the day and half of the second, and no *S* received reward on the same segment of the day's trials for more than 3 days consecutively.

Results and discussion.—The starting and two running times in the second runway, Trials 52–65, were transformed to reciprocals and subjected to analyses of variance. As a test for FE the data for the experimental groups were subjected to analyses in which the main effects were (*a*) delay duration, (*b*) first goal-box condition (reward vs. nonreward), and (*c*) days. For the second clock,[2] the only significant *F* was for the interaction between delay and first goal-box condition, $F(1, 14) = 7.99$, $p < .05$. This interaction is depicted in Fig. 1, which shows that the 5-sec. experimental *S*s ran faster following nonreward than following reward, $t(7) = 3.19$, $p < .01$, but that the first goal-box condition apparently had no effect on the 5-min. *S*s, $t(7) = .46$, $p > .05$.

The experimental and control groups were compared by means of two-way analyses of variance of mean reciprocal times, Trials 52–65 combined. These means were computed from the rewarded trials for each of the experimental *S*s and from all trials for each of the control *S*s. No significant experimental vs. control main effects or interactions were found, a result compatible with Wagner's failure to support the "response-depression" interpretation of FE.

In summary, the FE was demonstrated in the present study, as in previous studies, when *S*s were confined for a brief duration in the unbaited goal box. However, when the duration of confinement in the first goal box was 5 min., no significantly increased vigor of responding was found following nonreward. These data therefore support the Amsel-Roussel hypothesis that the FE is a transitory phenomenon.

REFERENCES

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

WAGNER, A. R. The role of reinforcement and non-reinforcement in an "apparent frustration effect." *J. exp. Psychol.*, 1959, **57**, 130–136.

(Received December 24, 1962)

---

[2] Equivalent results were obtained from analyses of the first clock (start) time with the exception that delay was also a significant main effect, $F(1, 14) = 38.04$, $p < .01$. This was due to the lower mean reciprocal start time of the 5-min. delay *S*s (1.56 vs. 4.55 for the 5-sec. delay *S*s). No analyses of the 1/10-sec. third clock yielded significant results.